



industriales
etsii

**Escuela Técnica
Superior
de Ingeniería
Industrial**

UNIVERSIDAD POLITÉCNICA DE CARTAGENA

Escuela Técnica Superior de Ingeniería Industrial

MONITORIZACIÓN DE LA EFICIENCIA DE DEPURACIÓN DE UNA PLANTA DE TRATAMIENTO DE AGUA RESIDUAL URBANA MEDIANTE EL EMPLEO DE TÉCNICAS DE ANÁLISIS ESTADÍSTICO MULTIVARIANTE

TRABAJO FIN DE MÁSTER

**MÁSTER EN INGENIERÍA AMBIENTAL Y DE PROCESOS
SOSTENIBLES**

Autor: JAVIER CARPE CÁNOVAS
Director: FRANCISCO JAVIER BAYO BERNAL
Codirector: JOAQUÍN LÓPEZ-CASTELLANOS GARCÍA
(AQUALOGY)



**Universidad
Politécnica
de Cartagena**

Cartagena, Septiembre de 2015

ÍNDICE

1. Resumen (Abstract)

2. Objetivos

3. Antecedentes bibliográficos

3.1. Introducción.

3.2. Descripción EDAR Cabezo Beaza.

3.3. Técnicas quimiométricas (PCA y HCA) aplicados a solucionar o evaluar problemas medioambientales.

3.4. Técnicas quimiométricas aplicadas a solucionar o evaluar problemas del agua.

4. Materiales y Métodos

4.1. Recolección y conjunto de datos muestrales.

4.2. Análisis estadístico de datos experimentales.

4.3. Análisis de datos multivariados.

5. Resultados y Discusión

5.1. Análisis principal de componentes.

5.1.1. Test K.M.O.

5.1.2. Comunalidades y grafico de sedimentación.

5.1.3. Matriz de componentes.

5.1.4. Análisis de factores principales.

5.2. Análisis de conglomerados jerárquico.

6. Conclusiones

7. Referencias bibliográficas

ÍNDICE DE FIGURAS

Figura 3.1: Localización del polígono industrial “Cabezo Beaza”.....	12
Figura 3.2: Localización de la EDAR “Cabezo Beaza”.....	12
Figura 3.3: Vista aérea de la EDAR.....	13
Figura3.4: Vista aérea de la EDAR (Fuente: http://www.inima.es/es/edar_cabezo_beaza_murcia).....	13
Figura 3.5: Diagrama de flujo de agua y fangos en la EDAR Cabezo Beaza de Cartagena.....	14
Figura 3.6: Esquema de la línea de agua de la EDAR (Fuente: Wikipedia).....	15
Figura 3.7: Esquema de la línea de fangos de la EDAR (Fuente: Wikipedia)...	16
Figura 5.1: Grafico de sedimentación.....	49
Figura 5.2: Variación de todos los componentes principales en función de las estaciones del año.....	55
Figura 5.3: Variación de todos los componentes principales en función de los días de la semana.....	56
Figura 5.4: PC5 representado en función de los meses del año.....	58
Figura 5.6: Barras de error de los PCs, en función de los días de las semana...	60.
Figura 5.7: Barras de error de los PCs, en función de las estaciones del año....	60
Figura 5.8: Barras de error del PC3 en función de los tres años de estudio.....	64
Figura 5.9: Dendograma con vinculación Ward.....	67
Figura 5.10: Clasificación de los resultados del clúster según valores autoescalados para los diferentes parámetros de calidad del agua.....	69

ÍNDICE DE TABLAS

Tabla 4.1: Análisis estadístico univariante.....	33
Tabla 5.1: Guía de interpretación para los resultados del test KMO.....	42
Tabla 5.2: Resultados del test KMO en nuestro ACP.....	43
Tabla 5.3: Comunalidades.....	45
Tabla 5.4: Varianza total explicada.....	48
Tabla 5.5: Matriz de componentes utilizando análisis de componentes principales como método de extracción.....	51
Tabla 5.6: Matriz de componentes utilizando análisis de componentes principales como método de extracción, eliminando de los componentes los pesos inferiores de los parámetros que estaban repetidos en otros componentes.....	53
Tabla 5.7: Prueba de homogeneidad de varianzas para los días de la semana...	62
Tabla 5.8: Prueba de homogeneidad de varianzas para las estaciones del año.....	63
Tabla 5.9: ANOVA del factor PC3 referido a los tres años de estudio.....	65

1. Resumen (Abstract)

En una gran cantidad de campos nos encontramos con un amplio número de variables que se miden a diario, por lo que se requiere un enfoque coherente y estructurado del sistema para entender su comportamiento inherente. En este sentido, son ampliamente utilizados tanto el análisis de componentes principales (ACP) como el análisis de conglomerados jerárquico (HCA), que son técnicas multivariantes utilizadas para extraer y estructurar la información.

En este trabajo, ambas herramientas estadísticas se aplicaran en una estación depuradora de aguas residuales urbanas (EDAR) situada en Cartagena, en el sureste de España, una zona marcada por un fondo de composición geoquímico muy característico del agua y un uso importante de fertilizantes en tareas agrícolas. Para ello, se tomaron muestras de agua durante tres años (2010, 2011 y 2012) y se analizaron una serie de parámetros como caudal, pH, conductividad, temperatura, MLVSS, sólidos en suspensión, V30, índice volumétrico de fangos, DBO₅, DQO, nitrógeno (amonio y nitrato) y fósforo (fosfato), en de tres puntos de la depuradora: el influente (salida del tratamiento primario hacia el secundario), el reactor biológico y el efluente. En total, se extrajeron cinco componentes en el análisis de componentes principales, que mostraron estar asociados con: los nutrientes y la materia orgánica a lo largo de todo el proceso (PC1), un componente iónico (PC2), la eficiencia del proceso de

depuración (PC3), la carga orgánica a la entrada de la EDAR (PC4) y la posible entrada de contaminantes a la depuradora procedentes de la erosión eólica (PC5). Se realizaron barras de error y pruebas de homogeneidad de varianzas (ANOVA) para los componentes principales con respecto a los días de la semana y las estaciones del año. Para los días de la semana, no aparecieron diferencias con significación estadística para los cuatro primeros componentes principales, aunque sí para el último, que indicaría que el PC5 está sujeto a cambios meteorológicos. En cuanto a las diferencias por estaciones, fueron significativas las variaciones para el componente principal 2 (PC2), que representaba la carga iónica de la EDAR, siendo más alto en verano que en el resto de estaciones, de forma estadísticamente significativa. El descenso de carga orgánica que se produce en la EDAR durante el estío, justificaría el incremento en este segundo componente, que representa la carga geoquímica de nuestras aguas. En cuanto al análisis de conglomerados jerárquicos, nos permitió distinguir los parámetros afluentes de los del efluente y en un análisis más exhaustivo pudimos comprobar que los resultados eran muy parecidos a los obtenidos en el análisis de componentes principales.

2. Objetivos

El **objetivo general** del presente Trabajo Fin de Máster es el de conocer y predecir la eficiencia en el proceso de depuración de una EDAR urbana mediante el uso de técnicas multivariantes. Para ello, nos planteamos los siguientes **objetivos específicos**:

- 1) Analizar los diferentes parámetros de medida de calidad del agua, tanto a la entrada como a la salida de la depuradora, para su uso como indicadores en los distintos componentes en los que se sintetizará toda la información disponible.
- 2) Decidir el número de componentes principales o factores a considerar en el estudio, de forma que se simplifique el manejo de demasiadas variables originales pero sin obviar información relevante de todo el proceso, que nos permita tomar decisiones.
- 3) Averiguar las relaciones de cada componente o factor con las variables originales, especificando la función que dicho componente tendrá en el modelo de componentes principales propuesto.

- 4) Emplear el análisis de conglomerados jerárquico como herramienta de análisis quimiométrico que aporte más información a partir de los datos y variables de partida.

- 5) Investigar las variaciones de los componentes principales propuestos con los días de la semana y las estaciones del año que nos permita encontrar un patrón de funcionamiento de la depuradora.

- 6) Razonar y discutir los resultados obtenidos y sacar conclusiones para la mejora de la planta.

3. Antecedentes bibliográficos

3.1 Introducción.

El agua ha sido, desde la creación del mundo, un elemento imprescindible para todos los seres vivos, necesaria para su mantenimiento, para formar parte de su composición o como medio para vivir. En la naturaleza podemos encontrar el agua con su calidad original unida a sustancias de origen natural, disueltas y en suspensión, orgánicas e inorgánicas, que va adquiriendo a lo largo de su recorrido. El desarrollo de las poblaciones unido a su actividad industrial, ha traído consigo un uso cada vez más creciente del agua, lo que ha llevado a una mayor explotación de los recursos hídricos y a un incremento excesivo de la contaminación del medio acuático. Esta pérdida de calidad en el agua, debida de manera directa o indirecta a la actividad humana, provoca en los países desarrollados la degradación del medio ambiente.

Para evitar estos efectos se han promulgado normas que limitan de manera cualitativa y cuantitativa los vertidos de las poblaciones y obliga al tratamiento de las aguas residuales urbanas mediante estaciones depuradoras donde reciben las aguas entregadas por los colectores de saneamiento (puntos finales de las redes de alcantarillado de la población). Un sistema de saneamiento integral convencional se componen de las redes de alcantarillado responsables de

recoger y transportar todas las aguas residuales y pluviales generadas en una población a una Estación Depuradora de Aguas Residuales (EDAR) que permite el tratamiento de las citadas aguas para la obtención de una determinada calidad en función de su uso posterior.

Una Estación Depuradora de Aguas Residuales (EDAR), también llamada planta de depuración o Planta de Tratamiento de Aguas Residuales (PTAR), consta de tres líneas de tratamiento. La más importante es la línea de agua, donde se elimina la mayor parte de la contaminación, tanto soluble como insoluble, que puede combinar diferentes procesos físicos, químicos y/o biológicos. La contaminación extraída es tratada en la segunda línea en orden de importancia que es la línea de fangos. Finalmente, como subproducto de la línea de fangos se produce biogás que es utilizado como combustible para el calentamiento del fango o para la cogeneración de energía eléctrica y térmica.

El rendimiento del proceso y operación de estaciones de tratamiento de aguas residuales es importante para que los parámetros se adecuen a los requisitos legales impuestos por la Unión Europea, para poder cumplir con las estrictas normas de descarga de efluentes (Directiva 91/271/CEE), teniendo especial cuidado con el nitrógeno y el fosforo, puesto que ambos son reconocidos como los dos nutrientes principales en la eutrofización acuática (Zuthi *et al.*, 2013).

Debido a la gran cantidad de variables que se miden cada día, se requiere un enfoque coherente y bien estructurado del sistema para entender el comportamiento y la eficiencia del proceso. En este sentido, tanto el análisis principal componentes (ACP) y el análisis de conglomerados jerárquico (HCA) son técnicas multivariantes que han sido ampliamente aplicados para extraer y estructurar la información para diferentes propósitos y en diferentes campos.

3.2 Descripción EDAR Cabezo Beaza

La EDAR “Cabezo Beaza” se localiza al este de la ciudad de Cartagena, en el polígono industrial anexo a esta ciudad. Las instalaciones depuran la mayor parte de las aguas residuales del término municipal de Cartagena, incluidas las del núcleo urbano y otros pequeños núcleos de población cercanos, sirviendo a un total de 140.000 habitantes equivalentes.



Figura 3.1: Localización del polígono industrial “Cabezo Beza” (Verde)
(Fuente: Google maps).



Figura 3.2: Localización de la EDAR “Cabezo Beza” (Fuente: Google maps).



Figura 3.3: Vista aérea de la EDAR (Fuente: Google maps).



Figura 3.4: Vista aérea de la EDAR (Fuente:

http://www.inima.es/es/edar_cabezo_beaza_murcia)

Esta planta consta de un sistema convencional de tratamiento primario y dos reactores de lodos activados. El reactor biológico aerobio de flujo pistón tiene un volumen neto total de 12.150 m^3 y consiste en un tanque anaeróbico para la pre-desnitrificación seguido por un tanque facultativo, con un volumen de 2.025

m³ y aireado para el licor mixto con retorno a las zonas anóxicas y facultativas.

En la **Figura 3.5**, se muestra un esquema del funcionamiento de la EDAR.

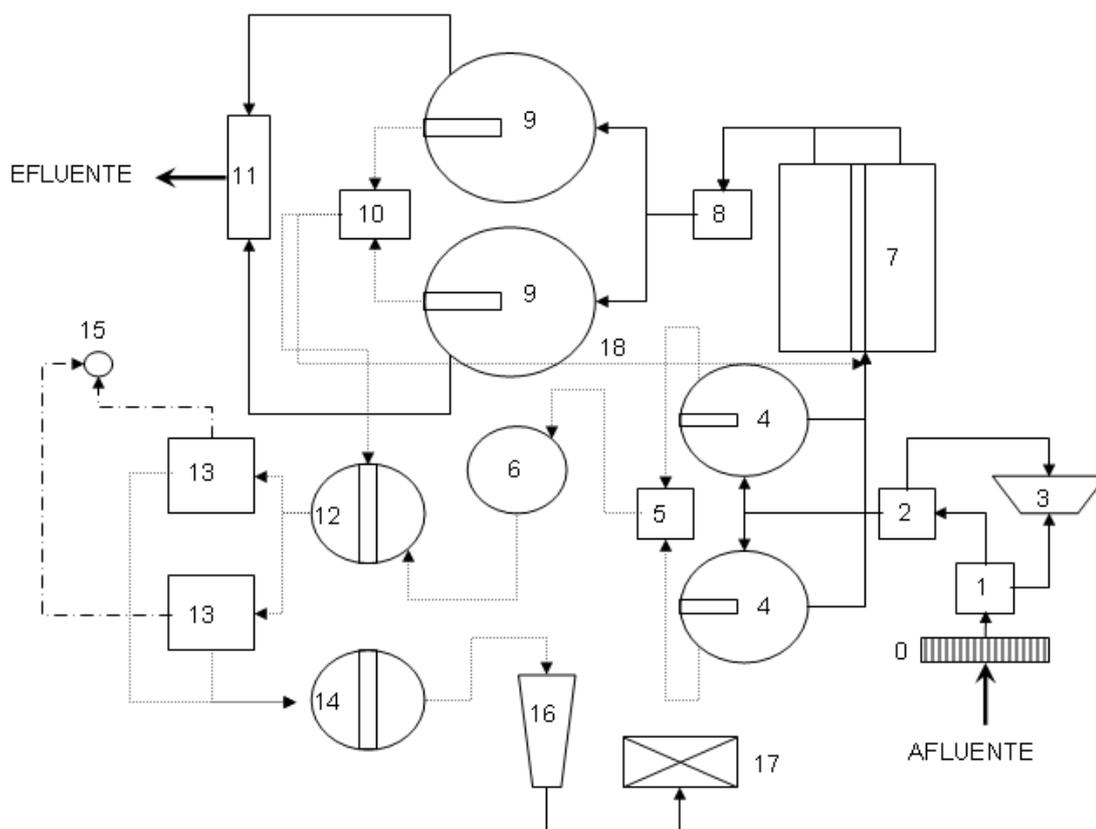


Figura 3.5: Diagrama de flujo de agua y fangos en la EDAR Cabezo Beza de Cartagena. 0. Rejas de desbaste; 1. Desarenado; 2. Eliminación de grasas y aceites; 3. Evacuación de residuos de pretratamiento; 4. Decantadores primarios; 5. Arqueta de mezcla de fangos primarios; 6. Espesamiento de fangos primarios; 7. Reactor biológico; 8. Arqueta de mezcla de fangos secundarios; 9. Decantadores secundarios; 10. Arqueta de mezcla de fangos secundarios decantados; 11. Cloración de agua depurada; 12. Espesamiento de fangos; 13. Digestores anaerobios; 14. Espesamiento de fangos digeridos; 15. Salida de gas a antorcha; 16. Deshidratación mediante centrifugas; 17. Evacuación de fangos; 18. Recirculación de fangos activos a reactor biológico

La línea de agua tiene como objetivo la reducción de la contaminación existente en el agua residual hasta los límites marcados en la legislación vigente. El primer proceso que nos encontramos es la obra de llegada y elevación donde se eleva el influente a una cota suficiente, para pasar por gravedad por los demás procesos unitarios de depuración hasta el punto de vertido al cauce receptor. Para ello consta de un sistema de elevación mediante bombeo que está protegido por una reja de pre-desbaste de amplia luz de paso. El siguiente proceso es el pretratamiento (mediante tratamientos físicos), a continuación nos encontramos con el tratamiento primario (por decantación primaria) y el último proceso de depuración es el tratamiento secundario (tratamiento biológico seguido de decantación secundaria). Todo ello seguido por un posible tratamiento terciario.

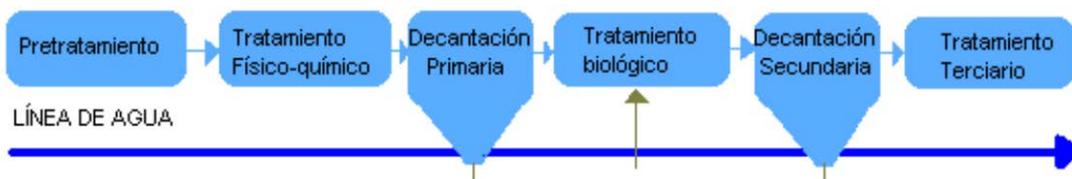


Figura 3.6: Esquema de la línea de agua de la EDAR (Fuente: Wikipedia).

La línea de fangos tiene como objetivo la retirada de los mismos que se han generado en la línea de agua. La finalidad de esta línea es reducir el volumen para abaratar los costes de transportes en su evacuación y minimizar en su

mayor parte la cantidad de microorganismos existentes en el fango al estabilizar los mismos.

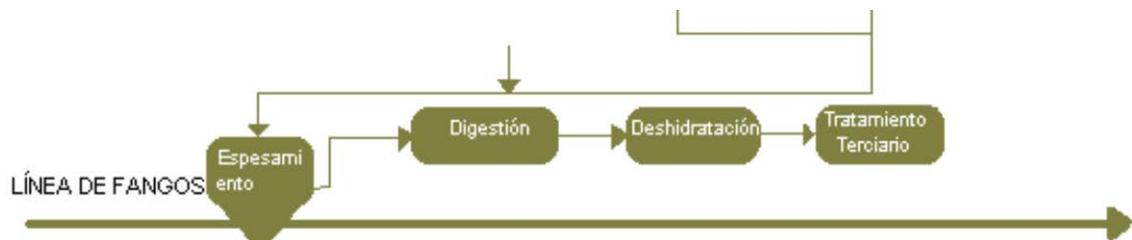


Figura 3.7: Esquema de la línea de fangos de la EDAR (Fuente: Wikipedia).

El tratamiento terciario se lleva a cabo a través de estanques de estabilización de baja carga terciarios o de maduración se diseñan para mejorar la calidad de los efluentes secundarios y para la nitrificación estacional. Los mecanismos biológicos involucrados son similares a los de otros procesos aerobios de cultivo en suspensión. Su funcionamiento implica la respiración endógena de los sólidos biológicos residuales y la conversión del amoníaco en nitrato mediante el oxígeno suministrado por reaireación superficial y por la presencia de algas. Se han propuesto tiempos de detención de 18 a 20 días como el mínimo periodo necesario para conseguir la respiración endógena completa de los sólidos residuales. Para mantener las condiciones aerobias, las cargas aplicadas deben ser bastante bajas.

3.3 Técnicas quimiométricas (PCA y HCA) aplicados a solucionar o evaluar problemas medioambientales.

Las técnicas quimiométricas son aquellas empleadas en la disciplina de la química que se enfoca en la aplicación de métodos matemáticos o estadísticos sobre datos químicos. Kiralj y Ferreira (2006) la definieron como: "...la disciplina química que utiliza métodos matemáticos y estadísticos para diseñar o seleccionar procedimientos de medida y experimentos óptimos, y para proporcionar la máxima información química mediante el análisis de datos químicos".

La investigación en quimiométrica abarca un amplio campo de métodos que pueden ser aplicados en química. Existen técnicas para la recolección correcta de datos (optimización de parámetros experimentales, diseño de experimentos, calibración, procesamiento de señales) y para obtener información a partir de estos (estadística, reconocimiento de patrones, modelado, etc.), estas últimas son las técnicas que nos interesan en este trabajo y utilizaremos como técnicas quimiométricas el ACP y el HCA (Xiaorui, 2000).

Los métodos multivariantes se aplican en la investigación científica con los siguientes objetivos:

1. Simplificación de la estructura inicial de los datos (reducción) sacrificando la mínima información posible.
2. Clasificación y agrupación de objetos o variables que son “similares” bajo un criterio o característica, de tal manera que sea más fácil la interpretación de la información una vez realizada la simplificación.
3. Relación entre las variables, mediante análisis de correlación.
4. Predicción a través de las relaciones que existan entre las variables.
5. Construcción de hipótesis y experiencias.

Estas técnicas se utilizan en estudios con grandes cantidades de datos cuyo manejo e interpretación pueden ser difíciles con técnicas estadísticas convencionales, esto ocurre, por ejemplo, en estudios y caracterizaciones de ciertos sistemas medioambientales. La naturaleza compleja que supone estudiar el medio ambiente requiere encontrar y proponer modelos simples para identificar y estudiar las variables que tengan un mayor impacto ambiental, por esta razón, el uso de las técnicas quimiométricas multivariantes será de gran ayuda para encontrar soluciones y conclusiones que nos ayuden a modelar potenciales riesgos ambientales. Hoy día las técnicas multivariantes no solo se

están aplicando a estudios medio ambientales, sino que se están aplicando en multitud de campos (medicina, psicología, medio ambiente, etc.). La inherente dificultad que supone hallar una explicación o una toma de decisión objetiva cuando se trabaja con muchas variables hace que se recurra a estos tipos de análisis.

El análisis de factores principales (ACP) se puede llevar a cabo para cuantificar la importancia de las diferentes variables en el conjunto de datos (Zhang *et al.*, 2010). Es una técnica estadística multivariante ampliamente utilizado en los estudios medioambientales para reducir los datos correlacionados en algunos componentes no correlacionados o factores que explican la varianza máxima (Vialle *et al.*, 2011). Se calcularon los componentes o factores principales (PCs) basado en una matriz de correlación.

Rodriguez *et al.* (2010) analizaron diferentes variedades de cebolla con un análisis Clúster y un análisis de componentes principales. Se estudiaron ocho variedades diferentes de cebolla de tres colores (blanco, dorado y rojo) que crecieron en las mismas condiciones ambientales y agronómicas. El análisis de conglomerados y el análisis de componentes principales no se agruparon de acuerdo a la variedad (grado de similitud genética, en base a diferentes fenoles totales y flavonoides), sino que fue el color la variable con mayor influencia.

Johar y Savage (2013) examinaron la información sobre una serie de limitaciones de la salud (enfermedades crónicas, tratamientos y consumos de drogas) para revelar la prevalencia y severidad de la insalubridad que no se observa de modo directo. Se utilizó el análisis de conglomerados que se aplicó a 265.468 personas que participaron en este estudio en Australia. Para los individuos más jóvenes, la insalubridad no viene dada por cuestiones físicas, sino por limitaciones mentales y para los individuos de mayor edad, las limitaciones estaban provocadas por la dependencia a 3 grupos de fármacos como mínimo, pero mostraban un mejor estado de salud mental. Para mejorar esta circunstancia se dedujo que se requería una asignación más eficaz de los recursos sanitarios.

Zibert y Praznikar (2012) hicieron un seguimiento de los componentes de contaminación del aire como la materia particulada (PM10) y el carbón negro (BC), que puede proporcionar información sobre la calidad del aire y la dinámica de las emisiones. La calidad del aire depende de las fuentes de emisiones naturales y antropogénicas, así como de las condiciones meteorológicas. Se tomaron muestras de BC y PM10 durante el periodo diurno de un año en el Puerto de Koper (Eslovenia) y se agruparon de acuerdo a la similitud del BC y el PM10 por hora, sin tener en cuenta, a priori, las condiciones meteorológicas, la temporada (fría o caliente) o si era día de semana o de fin de semana. Se utilizó un análisis clúster con la distancia euclídea al

cuadrado y el análisis mostro diez grupos para el caso del BC, 3 de estos grupos con influencia en un solo día y los 7 restantes mostraban perfiles similares abarcando más de un día. Resultados similares se encontraron para el caso del PM10, puesto que había un total de 8 grupos, de los cuales solo uno de ellos mostro un perfil que tenia influencia para un solo día, mientras que los 7 restantes presentaban perfiles para varios días. El análisis de conglomerados reveló que los patrones menos pronunciados y las concentraciones diarias y horarias bajas para el PM10 y BC se daban la mayoría de los días del año del análisis, pero se encontraron picos individuales pronunciados que pueden explicarse por perfiles de velocidades altas del viento. Se encontró un pico muy alto la medianoche del 30/04/2010 para los dos parámetros, esto se explica porque el 1 de mayo es la fiesta nacional de Eslovenia y hay una tradición muy arraigada de hacer hogueras. También se mostró una agrupación diurna con perfiles en diferentes días en el periodo de frio, mientras que en el periodo de calor no se mostraron estos perfiles, esto se explica con el uso de las chimeneas en la época fría. El análisis mostro que tanto las concentraciones de BV y las concentraciones de PM10 fueron incrementadas por la intensidad de la lluvia y velocidad del viento.

Peng *et al.* (2014) evaluaron la calidad del aroma de la pasta de soja tradicional china durante el almacenamiento, mediante un análisis de componentes principales. La pasta de soja, es un producto de soja fermentada y es

ampliamente utilizado para dar sabor en el Este y los países del sudeste asiático. El aroma característico de la pasta de soja es muy importante a lo largo de su vida útil, este estudio extrae volátiles compuestos a través de cromatografía de gases y un análisis con espectro de masas, logrando un análisis cuantitativo de 15 compuestos volátiles en la pasta de soja. En los 28 días de almacenamiento de la pasta de soja, los resultados mostraron que los compuestos volátiles alcohol y aldehído disminuyeron en un 35% y 26%, respectivamente. Por el contrario, el ácido, el éster y el compuesto heterocíclico se incrementaron en un 130%, 242% y 15%, respectivamente, esto implica que el aroma se transforma del floral inicial a un aroma más tostado. De acuerdo con la agrupación de la muestra en el análisis de componentes principales, la vida de almacenamiento de la pasta de soja se podría dividir en tres periodos, estos tres periodos representan el correspondiente al aroma floral, tostado y picante.

Lu *et al* (2011) investigaron el comportamiento de la densidad media del polietileno en el medio natural, utilizaron un análisis de componentes principales a partir de un conjunto de múltiples propiedades del polietileno para establecer una expresión Z de datos adimensionales. Se llevaron diferentes pruebas de exposiciones naturales del polietileno durante un año en diferentes periodos de prueba en las Islas Xisha en China. Se midieron propiedades de superficie y mecánicas del polietileno y se encontró que Z aumento gradualmente con el tiempo de exposición en combinación con los datos

climáticos en las islas Xisha. El efecto protector de los productos de oxidación generados por el proceso de degradación, fueron las condiciones que determinaron que el componente principal (tiempo de exposición y climatología) afectara al cambio de tendencia de Z. Por esto, Z puede ser utilizado para revelar el patrón de la degradación completa del polietileno en entornos naturales de una forma aproximada.

Baker (2010) realizó un análisis clúster de las vías de transporte aéreo de larga distancia y asociación de las concentraciones de contaminantes en el Reino Unido. El estudio se realizó en Birmingham en el periodo comprendido entre enero de 1998 a diciembre de 2001, se diferenciaron tres alturas para comprender mejor la influencia de la meteorología en la contaminación de esta región. Se identificaron seis patrones de transporte, mostrando variación estacional en la frecuencia, en las que se encontraron diferencias significativas en las concentraciones y el comportamiento de contaminantes atmosféricos cuando se analizaron las mediciones tomadas en una zona urbana y una zona rural. Las concentraciones más altas de contaminantes primarios se asociaron a masas de aire lento hacia el este del continente europeo, mientras que las concentraciones más bajas fueron asociadas con masas de aire del suroeste que pasan sobre el océano Atlántico. Se observó que había una mayor concentración de ozono en la temporada cálida y una concentración de ozono más baja en la

temporada fría, esto podría explicarse por la producción de ozono fotoquímico durante la estación cálida y NO_x (precursor del ozono troposférico).

3.4 Técnicas quimiométricas aplicadas a solucionar o evaluar problemas del agua.

Las técnicas quimiométricas también se utilizan en el estudio de la calidad del agua, y en especial de aquellas aguas destinadas al uso o consumo humano, en las que resulta imprescindible garantizar su buen estado y asegurar el bienestar de todas aquellas personas que vayan a hacer uso de ésta, así como las condiciones óptimas de la misma para mantener la biodiversidad de las especies que habitan su entorno.

Una clara muestra de este interés se refleja en las múltiples investigaciones que tienen como objetivo principal analizar la calidad del agua (Beamonte *et al.*, 2004, 2007; Shrestha y Kazama, 2006), definir y/o utilizar indicadores que permitan llevar a cabo dicho análisis (Beamonte *et al.*, 2004, 2010, 2012).

Karavoltzos *et al.* (2008) estudiaron la calidad del agua potable en varias regiones de Grecia, investigando parámetros físico-químicos como conductividad, sólidos disueltos totales y pH y otros parámetros relacionados con el agua potable como cloruros, potasio, calcio, magnesio y sodio, así como

metales pesados, cadmio, cobre, plomo, cromo y níquel. Los valores establecidos por la Directiva 98/83/CE fue superada por el plomo, cloruro y níquel. La mayoría de los problemas fueron encontrados en las zonas con escasez de recursos hídricos y la mayoría de los problemas relacionados con el agua potable era por una contaminación de las aguas subterráneas, tanto de origen antropogénico, como natural, así como la intrusión del agua de mar en los acuíferos.

Las técnicas quimiométricas se han utilizado en el campo del agua para investigar la compleja estructura que subyace a muchos de los datos analíticos, la reducción de los recursos de datos para algunos importantes variables, factores o componentes, sin perder información importante (Noori y cols., 2010). En particular, el ACP es una técnica de reducción de datos de gran alcance que se ha llevado a cabo para reducir una gran cantidad de variables de calidad del agua y los datos en un conjunto manejable de factores no correlacionados que explican la mayor parte de la variación observada, lo que nos permite generar hipótesis sobre los principales componentes identificados y su significado. El HCA complementa muy bien al ACP, la agrupación y el descubrimiento de estructuras intrínsecas latentes o comportamiento subyacente de los datos (Vialle *et al.*, 2011). Vialle *et al.* (2011) investigaron la calidad del agua de la escorrentía del techo de una casa para utilizarla como agua doméstica. Se midió la conductividad, aniones, color, dureza total, pH, entre otros

parámetros, además también se analizó el agua para *Escherichia coli*, coliformes totales y enterococos. Para saber qué factores o componentes afectaban a la composición del agua de lluvia se hizo un análisis de componentes principales y un análisis de conglomerados y los dos análisis diferenciaron tres grupos: composición iónica, parámetros relacionados con la carga microbiológica e indicadores de contaminación fecal. En general, el agua de lluvia recogida tenía buena calidad fisicoquímica, pero no cumplió con los requisitos para agua potable, además se detectó contaminación microbiológica del agua.

Olsen *et al*, (2012) hablan sobre la calidad del agua en grandes cuencas fluviales lo que supone una amplia colección de muestras de agua en el tiempo y en varios lugares diferentes dentro de la cuenca. El tamaño y la complejidad del conjunto de datos resultante hacen que el análisis global sea muy complicado, por lo que se tienen que emplear los métodos estadísticos multivariados, siendo el ACP una buena herramienta para reducir componentes y entender las relaciones entre ellos.

Zhang *et al*, (2010) investigaron la calidad del agua superficial en un río con intensa actividad humana en un conjunto de datos recogidos de 2004 a 2008. Para ello utilizó un análisis de conglomerados, que agrupó las 15 estaciones de muestreo diferentes en 5 grupos en las que la química del agua y los niveles de contaminación eran similares. Además, se realizó un análisis de componentes

principales, del cual se extrajeron cuatro componentes principales: nutrientes, metales pesados, componentes naturales y los componentes orgánicos, todo esto reveló la tendencia decreciente de los metales pesados, una tendencia creciente anual del componente orgánico y una tendencia estacional para el factor natural.

Algunos ejemplos más del uso de los métodos quimiométricos aplicados a aguas residuales son los siguientes: la composición de las aguas residuales de origen doméstico (Singh *et al.*, 2005), agua superficial (Hussain *et al.*, 2008;), calidad la distribución de los contaminantes orgánicos (Saba y Su, 2013; Zhang *et al.*, 2013), la distribución de metales pesados en las aguas (Bingöl *et al.*, 2013), los nuevos instrumentos de la política en sistemas de tratamiento de agua domestica (Naughton y Hynds, 2014), aguas minerales (Astel *et al.*, 2014) y plantas de tratamiento de aguas residuales (Aguado y Rosen, 2008; Ouali *et al.*, 2009), entre otros. Ouali *et al.*, 2009 presentaron un conjunto de datos descriptivos y analíticos cuantitativos de un estación depuradora de aguas residuales (EDAR) de Sfax, ubicado en el sur-este de Túnez que han sido procesados por técnicas estadísticas multivariantes para investigar la evolución de la calidad del agua residual durante un periodo de 12 años (1984-1996) a partir de parámetros físico-químicos y biológicos. El análisis de componentes principales mostro la existencia de tres PCs significativos que representan el 74% de la varianza. El primero fue asignado a agua con relativamente baja carga orgánica, mientras que el segundo y el tercero fueron asignados al agua con un promedio alto de carga

orgánica y cargas de mineral. Este estudio presenta la necesidad y la utilidad de las técnicas multivariantes para la evaluación e interpretación de grandes conjuntos de datos complejos con el fin de obtener una mejor información sobre la calidad del agua y el diseño de la red de monitoreo para la gestión eficaz de los recursos hídricos.

Astel *et al*, 2008, presentaron una aplicación de mapas auto organizantes (self organizing maps) y un análisis de componentes principales para analizar, modelar e interpretar una gran vigilancia de datos de calidad del agua superficial. La aplicación de métodos quimiométricos hizo posible descubrir patrones de calidad específicos de los parámetros químicos y biológicos utilizados para supervisar la calidad del agua (relación entre la temperatura y el agua, turbidez, dureza y bacterias coliformes), los impactos estacionales durante el largo periodo de observación y la relativa independencia de la ubicación espacial de los sitios de muestreo.

Platikanov *et al*, (2014) utilizó las técnicas quimiométricas de análisis de componentes principales (ACP) y regresión de mínimos cuadrados parciales para explorar y analizar los parámetros de calidad del agua en plantas de tratamiento de aguas residuales, los resultados que obtuvieron permitieron el descubrimiento de relaciones importantes entre los parámetros muestreados y su dependencia cíclica diaria.

Hussain *et al* (2008), evaluaron la calidad del agua superficial conectada a un proyecto de riego en Al-Fadhley al este de Arabia Saudi mediante un análisis de conglomerados. La idea de llevar a cabo este método era para utilizar todos los parámetros de la química del agua disponibles, incluyendo oligoelementos y otros iones que no se consideran en las técnicas convencionales para evaluaciones de calidad del agua. Se esperaba cierto grado de contaminación del agua, por pertenecer la zona a un proyecto de riego, en el que se utilizan fertilizantes, insecticidas y plaguicidas. Se tomaron 17 muestras tanto en verano como en invierno, para obtener un mayor conocimiento de la variación estacional de la calidad del agua y se analizaron las muestras recogidas para un total de 23 parámetros de calidad del agua incluyendo pH, conductividad, sodio, potasio, nitrato, arsénico, boro, cobalto, hierro, litio, manganeso, selenio, mercurio, entre otros. El análisis clúster dio lugar a tres tipos distintos de agua, tanto para las temporadas de verano como las de invierno, en las que había dos fuentes principales de contaminación en aguas superficiales y aguas subterráneas poco profundas, estas dos fuentes de contaminación son: Fertilizantes, micronutrientes, plaguicidas e insecticidas, utilizadas en las actividades agrícolas, y las fuentes naturales no puntuales.

Gazzaz *et al*, (2012) estudió la caracterización de los patrones espaciales de la calidad del agua del río utilizando técnicas quimiométricas. Este estudio empleó

tres técnicas quimiométricas: Análisis de componentes, análisis de conglomerados y análisis discriminante para identificar la estructura latente de la calidad del agua de un río en Malasia, estas técnicas destacan por su gran potencial de reducir razonablemente el número de parámetros de calidad y estaciones de monitoreo, sin perder mucha información.

Con todos estos trabajos mostrados anteriormente, podemos asegurar que las técnicas quimiométricas son ampliamente utilizadas en un gran número de estudios en diferentes campos.

.

4 Materiales y Métodos

4.1 Recolección y conjunto de datos muestrales.

Se tomaron muestras de aguas residuales para diferentes parámetros en los años 2010, 2011 y 2012, midiéndose un máximo de 1.096 muestra para un parámetro y un mínimo de 155, desde el influente (después del tratamiento primario), el reactor biológico, la recirculación y el efluente en la estación de tratamiento de aguas residuales urbanas (EDAR) “Cabezo Beaza” situado en Cartagena (España). Los siguientes parámetros de calidad del agua se incluyeron en el conjunto de datos:

Flujo de agua [Qa] ($\text{m}^3 \text{d}^{-1}$), pH [pH](unidades de pH), la temperatura del agua [Ta] ($^{\circ}\text{C}$), conductividad eléctrica [CE] (mS cm^{-1}), sólidos en suspensión [SS] (mg l^{-1}), sólidos suspendidos volátiles licor mixto [MLVSS] (mg l^{-1}), el volumen del lodo sedimentado después de 30 minutos de sedimentación [V30] (ml l^{-1}), índice de volumen de fangos [IVF] (ml g^{-1}), la demanda química de oxígeno [DQO] (mg l^{-1}), la demanda bioquímica de oxígeno [DBO] (mg l^{-1}), nitrógeno amonio [$\text{NH}_4\text{-N}$] (mg l^{-1}), nitrógeno nítrico [$\text{NO}_3\text{-N}$] (mg l^{-1}), fósforo fosfato [$\text{PO}_4\text{-P}$] (mg l^{-1}). Se utilizan acrónimos con subíndices para indicar el efluente correspondiente: para el influente (I), reactor biológico (RB) o efluente (E) de la ETAR. La CE se determinó con un conductrimetro Crison GLP 32 (Barcelona,

España) y el pH con un pH-metro Crison GLP 22 (Barcelona, España), calibrado por medio de soluciones estándar. La DQO se determinó utilizando el Spectroquant NOVA 30 de Merck (Darmstadt, Alemania) y se utilizaron respirómetros oxytop (WTW, Weiheim, Alemania) para el análisis de DBO. Todos los demás parámetros se calcularon de acuerdo con los Métodos Estándares para el Análisis de Aguas y Aguas Residuales (APHA, 2012).

4.2 Análisis estadístico de datos experimentales.

El tratamiento estadístico de los datos se realizó con el programa SPSS (Paquete Estadístico para las Ciencias Sociales) 22.0. La **Tabla 4.1** muestra el mínimo, máximo, media, la mediana y la desviación estándar correspondiente a los parámetros medidos en la planta de tratamiento de aguas residuales, evitando valores por debajo del límite de cuantificación para cada técnica analítica. Además, representa el número de valores perdidos, puesto que solo los caudales fueron medidos a diario, siendo estos los únicos con 0 valores perdidos. El método de Tukey-Kramer fue el procedimiento de comparaciones múltiples que se utiliza para la estimación simultánea de diferencias por pares de medios en un modelo lineal, según lo recomendado por Storline (1981).

Tabla 4.1: Análisis estadístico univariante

	N		Media	Mediana	Desv. típ.	Mínimo	Máximo
	Válidos	Perdidos					
[Qa]i (m ³ /d ⁻¹)	1096	0	23598,2	23425,5	4356,36	4159	47011
[Qa]rb (m ³ /d ⁻¹)	1096	0	24628,8	24104	4293,81	3000	50000
[Qa]e (m ³ /d ⁻¹)	1096	0	21881,5	22000	3472,09	3784	39151
[pH]i (unidad pH)	157	939	7,68	7,71	0,172	6,93	8,11
[pH]rb(unidad pH)	678	418	6,97	6,97	0,227	4,11	7,46
[pH]e (unidad pH)	487	609	7,57	7,61	0,304	6,61	8,34
[CE]i (mS cm ⁻¹)	158	938	2,15	2,18	0,301	1,14	2,93
[CE]rb (mS cm ⁻¹)	678	418	1,91	1,94	0,432	0,82	9,96
[CE]e (mS cm ⁻¹)	487	609	1,96	1,98	0,307	0,91	2,87
[DQO]i (mg l ⁻¹)	157	939	457,77	469,00	92,04	186,00	662,00
[DQO]e (mg l ⁻¹)	482	614	39,63	37,45	13,63	8,30	117,00
[DBO]i (mg l ⁻¹)	157	939	279,39	290,00	68,81	30,00	440,00
[DBO]e (mg l ⁻¹)	212	884	7,83	7,00	4,832	1,30	50,00
[P-PO ₄]i (mg l ⁻¹)	155	941	15,49	16,60	6,91	0,80	43,60
[P-PO ₄]e (mg l ⁻¹)	228	868	1,68	0,80	2,217	0,08	12,60
[SS]i (mg l ⁻¹)	157	939	0,28	0,20	0,282	0,00	1,50
[V30]rb (ml l ⁻¹)	678	418	514,60	447,50	241,18	110,00	965,00
[SS]e turbidez	474	622	4,61	4,10	3,245	0,60	42,00
[N-NH ₄]i (mg l ⁻¹)	155	941	50,70	50,20	12,95	19,20	92,20
[N-NH ₄]e (mg l ⁻¹)	222	874	18,91	12,70	17,51	0,02	61,80
[N-NO ₃]i (mg l ⁻¹)	155	941	0,80	0,70	0,635	0,20	5,70
[N-NO ₃]e (mg l ⁻¹)	230	866	6,02	5,20	4,652	0,00	27,00
[Ta]rb (°C)	678	418	23,39	23,45	3,882	15,20	32,25
[IVF]rb (ml g ⁻¹)	678	418	196,01	181,00	74,84	44,50	419,00
[MLSS]rb (mg l ⁻¹)	678	418	2560,36	2580,00	455,84	1302,00	3680,00
[MLVSS]rb (mg l ⁻¹)	676	420	1932,04	1959,90	337,05	823,80	2736,30

4.3 Análisis de datos multivariados

El análisis multivariante es la rama de la estadística y del análisis de datos, que estudia, interpreta y elabora el material estadístico sobre un conjunto de $n > 1$ de variables, que pueden ser cuantitativas, cualitativas o una mezcla. Su razón de ser radica en un mejor entendimiento del fenómeno objeto de estudio obteniendo información que los métodos estadísticos univariantes y bivariantes son incapaces de conseguir (Cuadras, 1981). Se pueden clasificar en tres grandes grupos:

- 1) Métodos de dependencia: Suponen que las variables analizadas están divididas en dos grupos, las variables dependientes y las variables independientes. El objetivo de los métodos de dependencia consiste en determinar si el conjunto de variables independientes afecta al conjunto de variables dependientes y de qué forma. Se pueden clasificar en dos grandes subgrupos, según que la variable dependiente sea cuantitativa o cualitativa. Si la variable dependiente es cuantitativa, algunas de las técnicas que se pueden aplicar son las siguientes: Análisis de regresión, análisis de supervivencia, análisis de la varianza, correlación canónica. Si la variable dependiente es cualitativa, las técnicas que se pueden aplicar son las siguientes: Análisis discriminante, modelos de regresión logística, análisis de conjunto.

- 2) Métodos de interdependencia: Estos métodos no distinguen entre variables dependientes e independientes y su objetivo consiste en identificar que variables están relacionadas, como lo están y porque. En este grupo de métodos se encuentran el análisis de componentes, el análisis clúster y el análisis de escalamiento
- 3) Métodos estructurales: Suponen que las variables están divididas en dos grupos: el de las variables dependientes y el de las independientes. El objetivo de estos métodos es analizar, no solo como las variables independientes afecta a las variables dependientes, sino también como están relacionadas las variables de dos grupos entre sí.

El Análisis de Componentes principales (ACP) es una técnica estadística de síntesis de la información, o reducción de la dimensión (numero de variables). Es decir, ante un banco de datos con muchas variables, el objetivo será reducirlas a un número menor, perdiendo la menor cantidad de información posible. Los nuevos componentes principales o factores serán una combinación lineal de las variables originales, y además serán independientes entre sí. Un aspecto clave en ACP es la interpretación de los factores, ya que esta no viene dada a priori, sino que será deducida tras observar la relación de los factores con las variables iniciales (habrá, pues, que estudiar tanto el signo como la magnitud de las correlaciones). Esto no siempre es fácil, y será de vital importancia el

conocimiento que el experto tenga sobre la materia de investigación. (Terrádez, M. 2003).

El ACP construye una transformación lineal que escoge un nuevo sistema de coordenadas para el conjunto original de datos en el cual la varianza de mayor tamaño del conjunto de datos es capturada en el primer eje (llamado Primer Componente Principal), la segunda varianza más grande es el segundo eje, y así sucesivamente. Para construir esta transformación lineal debe construirse primero la matriz de covarianza o matriz de coeficientes de correlación. Debido a la simetría de esta matriz existe una base completa de vectores propios de la misma. La transformación que lleva de las antiguas coordenadas a las coordenadas de la nueva base es precisamente la transformación lineal necesaria para reducir la dimensionalidad de datos. Además las coordenadas en la nueva base dan la composición en factores subyacentes de los datos iniciales. (Rescala, C. 2012).

En nuestro caso, utilizamos la rotación ortogonal varimax con normalización Kaiser como método de rotación, preferible a los métodos de rotación oblicua que son más complejos (Reimann *et al.*, 2002). Los componentes o factores principales (PCs) que obtuvimos en este trabajo fueron variables no correlacionadas ortogonales, obtenidas multiplicando las variables correlacionadas originales con el vector propio, siendo así combinaciones

lineales ponderadas de los originales (Singh *et al.*, 2005). Los PCs fueron exhibidos en el orden de variación descendente. Entre los criterios utilizados para probar la calidad de los resultados son (Reimann *et al.*, 2002, Gazzaz *et al.*, 2012):

- 1.- La varianza explicada de cada componente individual ($> 5\%$).
- 2.- La proporción total de la varianza acumulada explicada debe ser próxima al 70%.
- 3.- Una estructura simple de la matriz rotada que permitiría una interpretación más fácil de los resultados.

Si hay una alta correlación entre muchas de las variables originales, un gran porcentaje de la varianza total se explicaría por sólo unos pocos PCs. Se utilizó un estudio Kaiser-Meyer-Olkin (KMO) como prueba de adecuación muestral para estudiar el grado de ajuste de los datos a un análisis factorial, y la prueba de esfericidad de Barlett para examinar la hipótesis nula de que la matriz de correlación resultante es una matriz de identidad.

El Análisis Clúster, conocido como Análisis de Conglomerados, es una técnica estadística multivariante que busca agrupar elementos (o variables) tratando de lograr la máxima homogeneidad en cada grupo y la mayor diferencia entre los grupos. El Análisis Clúster tiene una importante tradición de aplicación en

muchas áreas de investigación. Sin embargo, junto con los beneficios del análisis clúster existen algunos inconvenientes. El análisis clúster es una técnica descriptiva, atórica y no inferencial. El Análisis Clúster no tiene bases estadísticas sobre las que deducir inferencias estadísticas para una población a partir de una muestra, es un método basado en criterios geométricos y se utiliza fundamentalmente como una técnica exploratoria, descriptiva pero no explicativa. Las soluciones no son únicas, en la medida en que la pertenencia al conglomerado para cualquier número de soluciones depende de muchos elementos del procedimiento elegido (De la Fuente 2011). Por otra parte, la solución clúster depende totalmente de las variables utilizadas, la adición o destrucción de variables relevantes puede tener un impacto substancial sobre la solución resultante. Los algoritmos de formación de conglomerados se agrupan en dos categorías:

- 1.- Algoritmos de partición: Método de dividir el conjunto de observaciones en k conglomerados (clúster), en donde “k” lo define inicialmente el usuario.
- 2.- Algoritmos jerárquicos: Método que entrega una jerarquía de divisiones del conjunto de elementos en conglomerados.

Un método jerárquico aglomerativo parte con una situación en que cada observación forma un conglomerado y en sucesivos pasos se van uniendo, hasta

que finalmente todas las situaciones están en un único conglomerado. Otro método jerárquico es el disociativo que sigue el sentido inverso, parte de un gran conglomerado y en pasos sucesivos se va dividiendo hasta que cada observación queda en un conglomerado distinto (Cuadras 2014).

Se utilizó el análisis de agrupamiento jerárquico por categoría de división en este trabajo, ya que los datos se tomaron como un clúster y secuencialmente se dividen en un dendrograma (o diagrama de árbol) de grupos, comenzando con el par más similar de objetos y la formación de grupos (Wunderlin *et al.*, 2001). La distancia de ligamiento se calculó como el cociente entre la distancia de ligamiento para un caso particular dividido por la distancia de ligamiento máxima (D_{link} / D_{max}), multiplicado por 100, como una forma de datos de estandarizar la distancia vinculación representado en el eje Y (Wunderlin *et al.*, 2001).

5 Resultados y Discusión

5.1 Análisis principal de componentes.

El análisis de componentes principales se llevó a cabo en 23 variables de las 26 variables indicadas anteriormente en la **Tabla 4.1**, excluyendo de los parámetros [Qa]i, [Qa]rb y [Qa]e. Todos estos atributos cumplieron el umbral de carga cruzada 0,4 en la matriz girada Varimax (**Tabla 5.5**), con una prueba anterior de Kaiser-Meyer-Olkin (**Tabla 5.2**), para ver si podríamos realizar este análisis ACP con garantías.

5.1.1 Test K.M.O

Lo primero que obtenemos es el test KMO (Kaiser-Meyer-Olkin), que tiene en cuenta las correlaciones para examinar la idoneidad de los datos del ACP. Gazzaz *et al* (2012) propone una guía de interpretación para los resultados del test KMO, la cual se muestra en la **Tabla 5.1**, procedente de Kaiser (1974):

Tabla 5.1: Guía de interpretación para los resultados del test KMO.

Valor KMO	Interpretación
0,90-1,00	Estupendo (Marvelous)
0,80-0,89	Meritorio (Meritorious)
0,70-0,79	Medio (Middling)
0,60-0,69	Mediocre (Mediocre)
0,50-0,59	Malo (Miserable)
0,00-0,49	Inaceptable (Unacceptable)

Para que el ACP pueda realizarse con garantías, se deben obtener valores por encima del 0,60. En nuestro caso, obtenemos una medida de adecuación muestral KMO de 0,70 (medio), por lo que podríamos aceptar este análisis. En la **Tabla 5.2**, se presentan estos resultados:

Tabla 5.2: Resultados del test KMO en nuestro ACP.

KMO y prueba de Bartlett		
Medida de adecuación muestral de Kaiser-Meyer-Olkin.		0,700
Prueba de esfericidad de Bartlett	Chi-cuadrado aproximado	1862,038
	Gl	253
	Sig.	0,000

Gazzaz *et al.* (2012) aportó un valor similar para el coeficiente KMO en el estudio los patrones espaciales en el agua de río, lo que indica un conjunto de datos susceptibles al análisis de componente principal. El criterio de selección de los componentes principales significativos (PCs) era elegir aquellos con valores propios o autovalores mayores que uno, también conocido como la regla de Kaiser (Kaiser, 1960). Por otra parte, la prueba de esfericidad de Bartlett fue significativa [$\chi^2 = 1862,038$, $p < 0,001$], el rechazo de la hipótesis nula de que la matriz de correlaciones es una matriz de identidad, y la aceptación de una clara correlación entre las categorías de respuesta y la pertinencia del modelo seleccionado, con menos datos que el original (Salvati y Carlucci, 2014).

5.1.2 Comunalidades y Grafico de sedimentación.

Las comunalidades representan la varianza de cada variable explicada por los componentes o factores principales. La comunalidad de una variable es la suma de las cargas factoriales asociadas a ella elevadas al cuadrado. Su computo se realiza a partir de la matriz de cargas factoriales. En el caso de componentes principales, cuando se retienen todas las variables, la comunalidad es siempre 1. Cuando se emplean otros métodos de extracción, la comunalidad inicial es el coeficiente de correlación múltiple entre cada variable y todas las demás.

La comunidad final suele tener mayor interés puesto que indica la variabilidad de cada variable que es explicada por el conjunto de factores retenidos. La columna de extracción corresponde a un porcentaje de la varianza de cada una de las variables. Esto se puede observar en la **Tabla 5.3**.

Tabla 5.3: Comunalidades

	Inicial	Extracción
[pH]i	1,000	0,575
[pH]rb	1,000	0,729
[pH]e	1,000	0,698
[CE]i	1,000	0,883
[CE]rb	1,000	0,956
[CE]e	1,000	0,898
[DQO]i	1,000	0,870
[DQO]e	1,000	0,841
[DBO]i	1,000	0,890
[DBO]e	1,000	0,718
[P-PO ₄]i	1,000	0,802
[P-PO ₄]e	1,000	0,685
[SS]i	1,000	0,713
[V30]rb	1,000	0,931
[SS]e	1,000	0,831
turbidez		
[N-NH ₄]i	1,000	0,674
[N-NH ₄]e	1,000	0,790
[N-NO ₃]i	1,000	0,670
[N-NO ₃]e	1,000	0,521
[Ta]rb	1,000	0,820
[IVF]rb	1,000	0,905
[MLSS]rb	1,000	0,855
[MLVSS]rb	1,000	0,854

La elección del número de componentes a considerar en el modelo puede hacerse según diversos criterios. Una aproximación razonable consiste en considerar el significado de los autovalores o valores propios, indicadores de la proporción de varianza explicada. La **Tabla 5.4** muestra los componentes extraídos, en la que los siete primeros factores extraídos, con valores propios (autovalores) mayores a uno y describiendo un 78,730%. Cada factor representa una variable en el modelo, aunque se compone de una combinación lineal de diferentes atributos. El primer componente (PC1) representó el 26,165% de la varianza total, el segundo componente (PC2), tercero (PC3), cuarto (PC4), quinto (PC5), sexto (PC6) y séptimo (PC7) representaron respectivamente el 16,071%, 10,945%, 8,089%, 7,032%, 5,646% y 4,781%. Esta generalizado el admitir que han de seleccionarse aquellos factores cuyo autovalor sea superior a 1, en nuestro caso, solo tienen autovalores mayores a 1 los siete primeros componentes. Obviamente esto conlleva una pérdida de capacidad explicativa, lo importante es que esa pérdida no sea significativa.

Obtenemos una varianza total (suma de valores propios o autovalores) de 22,99 ($6,018+3,696+2,517+1,861+1,617+1,299+1,100\dots\dots= 22,99$), con esta varianza total podemos calcular el porcentaje de varianza que explica cada uno de los factores principales. Por ejemplo, para el PC1 el porcentaje de varianza explicada se consigue sacando el tanto por cien a partir del total de los autovalores iniciales, en este caso 6,018 y la varianza total anteriormente

calculada, 22,99. De esta manera obtenemos el 26,165 % que explica el PC1
(Tabla 5.4).

Tabla 5.4: Varianza total explicada

Componente	Autovalores iniciales			Sumas de las saturaciones al cuadrado de la extracción			Suma de las saturaciones al cuadrado de la rotación		
	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado	Total	% de la varianza	% acumulado
1	6,018	26,165	26,165	6,018	26,165	26,165	3,653	15,881	15,881
2	3,696	16,071	42,236	3,696	16,071	42,236	3,023	13,143	29,024
3	2,517	10,945	53,181	2,517	10,945	53,181	3,017	13,117	42,141
4	1,861	8,089	61,270	1,861	8,089	61,270	2,680	11,653	53,793
5	1,617	7,032	68,303	1,617	7,032	68,303	2,465	10,716	64,509
6	1,299	5,646	73,949	1,299	5,646	73,949	1,906	8,286	72,795
7	1,100	4,781	78,730	1,100	4,781	78,730	1,365	5,935	78,730
8	0,959	4,171	82,900						
9	0,736	3,199	86,099						
10	0,616	2,678	88,778						
11	0,525	2,284	91,062						
12	0,421	1,832	92,894						
13	0,362	1,574	94,468						
14	0,277	1,206	95,674						
15	0,225	0,979	96,654						
16	0,192	0,835	97,489						
17	0,172	0,747	98,235						
18	0,158	0,685	98,920						
19	0,103	0,449	99,369						
20	0,092	0,401	99,770						
21	0,036	0,158	99,927						
22	0,014	0,060	99,988						
23	0,003	0,012	100,000						

Un método alternativo para la selección del número de componentes es el gráfico de sedimentación en el que se representan gráficamente los autovalores (eje de ordenadas) para cada factor (eje de abscisas). El análisis visual de este gráfico se centra en el punto de inflexión en la **Figura 5.1**, es habitual que los factores que estén por encima de uno se tomen como validos. Los autovalores del PC5, PC6 y PC7 son 1,6, 1,3 y 1,1 respectivamente, pero como explicaremos más adelante, no tomaremos como validos el PC6 y el PC7.

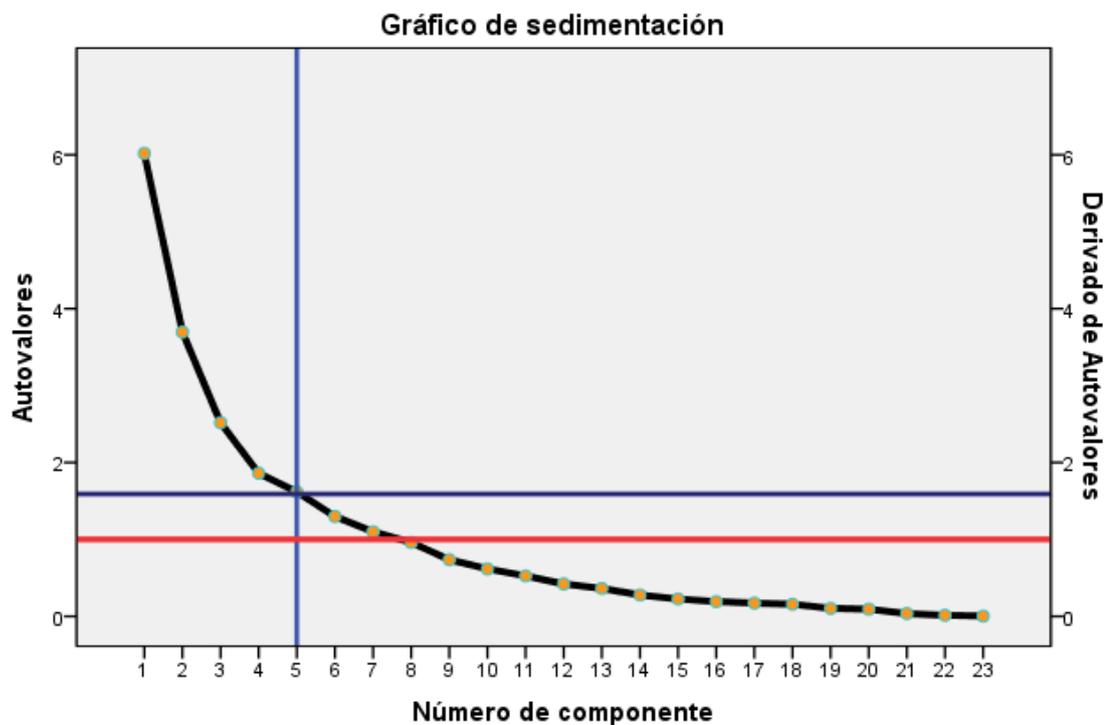


Figura 5.1: Gráfico de sedimentación. Línea roja horizontal representa el valor propio 1,1, correspondiente al PC7 y la línea azul oscuro horizontal representa el valor propio 1,6, correspondiente al del PC5. La línea azul clara vertical corresponde al número de componentes con el que nos quedamos.

5.1.3 Matriz de componentes.

Lo siguiente que encontramos es la matriz de componentes en la que aparecen los pesos de cada variable en cada uno de los factores. En total, obtenemos 7 componentes o factores principales y solo se muestran los pesos superiores a 0,4, para simplificar la matriz, esta condición la establecimos según Gazzaz *et al.*(2012).

Tabla 5.5: Matriz de componentes utilizando ACP como método de extracción

	Componente						
	1	2	3	4	5	6	7
[pH]i			0,467		0,444		
[pH]rb	0,563						
[pH]e	0,513	0,443					
[CE]i		0,834			-0,409		
[CE]rb		0,882					
[CE]e		0,791			-0,428		
[DQO]i	0,585			-0,596			
[DQO]e			0,776				
[DBO]i	0,572			-0,581			
[DBO]e	-0,410		0,601				
[P-PO ₄]i	0,654						
[P-PO ₄]e	-0,539			0,499			
[SS]i					0,519		
[V30]rb	0,734	-0,405					
[SS]e turbidez			0,692	0,421			
[N-NH ₄]i	0,585						0,552
[N-NH ₄]e	0,819						
[N-NO ₃]i					0,468		-0,509
[N-NO ₃]e	-0,619						
[Ta]rb	-0,476	0,590					
[IVF]rb	0,584	-0,496				-0,408	
[MLSS]rb	0,744						
[MLVSS]rb	0,780						

Según algunos estudios como en el de Gazzaz *et al.* (2012), si un mismo parámetro está repetido en uno o varios componentes principales solo tomamos el de mayor valor, descartando así, los demás. Por este motivo, solo aceptamos los 5 primeros factores o componentes principales, puesto que los parámetros presentes en el PC6 y PC7 están también presentes con un peso mayor en otros componentes principales. Una decisión similar se tomó Gazzaz *et al.* (20012), puesto que la exclusión de estos dos factores significó la reducción de la varianza explicada del 78,73% al 68,30% del conjunto de datos. Zhang *et al.* (2010) aportó una varianza explicada del 58,79% para los cuatro factores de control de la calidad del agua del río Xiangjiang, y Vialle *et al.* (2011) explicó un 60,00% de la varianza total para el monitoreo de la calidad del agua de escorrentía de un tejado. La **Tabla 5.5** quedaría representada de la siguiente manera en la **Tabla 5.6**:

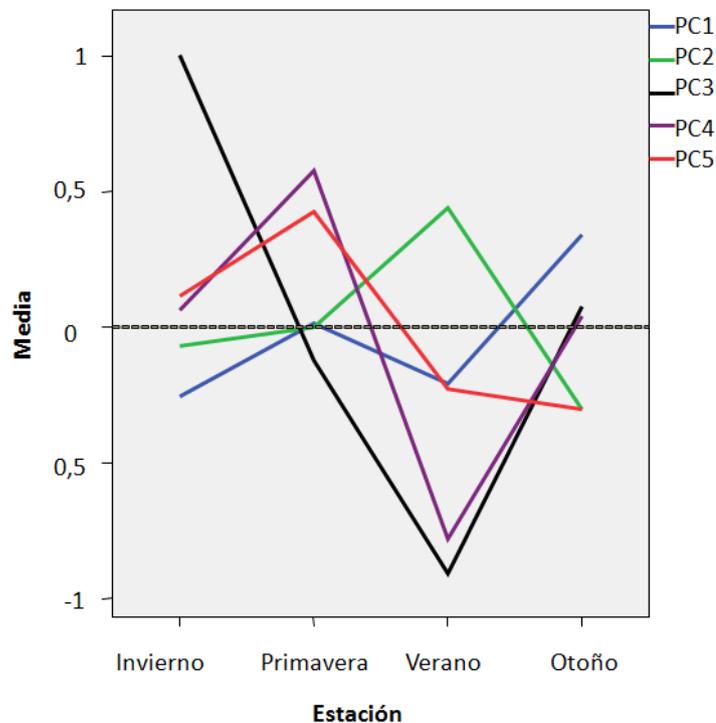
Tabla 5.6: Matriz de componentes utilizando análisis de componentes principales como método de extracción, eliminando de los componentes los pesos inferiores de los parámetros que estaban repetidos en otros componentes.

	Componente				
	1	2	3	4	5
[pH]i			0,467		
[pH]rb	0,563				
[pH]e	0,513				
[CE]i		0,834			
[CE]rb		0,882			
[CE]e		0,791			
[DQO]i				-0,596	
[DQO]e			0,776		
[DBO]i				-0,581	
[DBO]e			0,601		
[P-PO ₄]i	0,654				
[P-PO ₄]e	-0,539				
[SS]i					0,519
[V30]rb	0,734				
[SS]e turbidez			0,692		
[N-NH ₄]i	0,585				
[N-NH ₄]e	0,819				
[N-NO ₃]i					0,468
[N-NO ₃]e	-0,619				
[Ta]rb		0,590			
[IVF]rb	0,584				
[MLSS]rb	0,744				
[MLVSS]rb	0,780				

5.1.4 Análisis de factores principales.

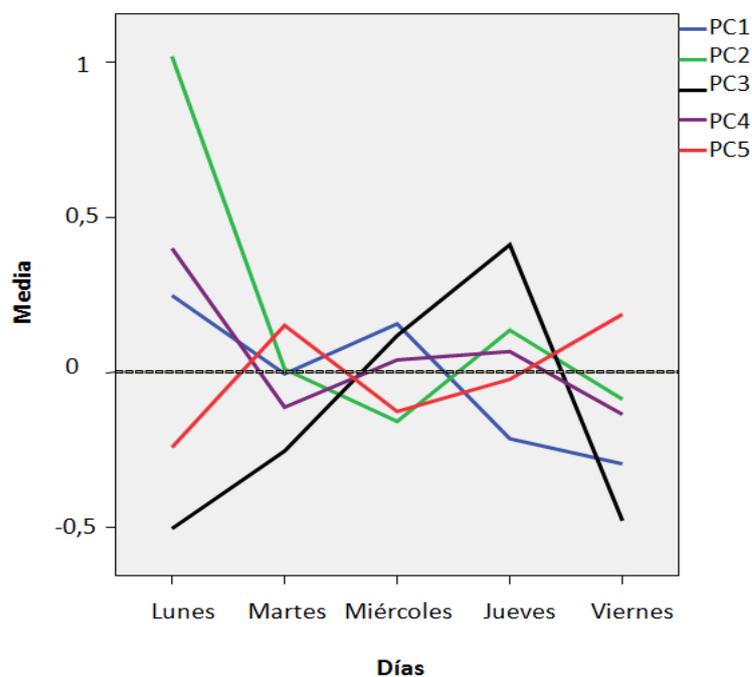
El primer componente o factor principal (PC1) explica un 26,16% de la varianza total y se caracteriza por tener el mayor número de atributos que cargan en un solo componente (once). Los parámetros con valores más altos son: $[\text{N-NH}_4]_e$ (0,819), $[\text{P-PO}_4]_i$ (0,654), $[\text{N-NO}_3]_e$ (-0,619), $[\text{N-NH}_4]_i$ (0,585) y $[\text{P-PO}_4]_e$ (-0,539), esto correspondería a una contaminación por nutrientes, sin embargo, también podemos observar valores altos en otros parámetros como son el $[\text{MLVSS}]_i$ (0,780) y $[\text{MLSS}]_{rb}$ (0,744), que explicaría la materia orgánica presente en el reactor biológico. Por lo que podríamos decir, que el primer componente o PC1 correspondería a la parte de los nutrientes y materia orgánica presentes en el proceso. Este componente principal está fuertemente influenciado con la contaminación por nutrientes procedentes de fuentes agrícolas, porque se demostró que aumentaba durante otoño y primavera, como se muestra en la **Gráfica 5.2**. De acuerdo con Acosta *et al.* (2011), estas son las estaciones en que las aplicaciones de fertilizantes y desechos del ganado en suelo ocurren en la Región de Murcia. Hussain *et al.* (2008) ha informado de un periodo similar para el riego y el uso de fertilizantes en Arabia Saudí.

Figura 5.2: Variación de todos los componentes principales en función de las estaciones del año.



El componente principal 1 (PC1), además, muestra valores positivos lunes, martes y miércoles, coincidiendo con un aumento en la actividad agrícola e industrial, y una reducción para el jueves y viernes. Según lo informado por otros investigadores (Beirle *et al*, 2003; Schipa *et al*, 2009), la actividad humana en los países industrializados sigue en gran parte un ciclo de siete días, con una disminución bien conocida de los contaminantes durante los fines de semana, en el llamado “efecto de fin de semana”. Esto se puede observar en la **Figura 5.3**.

Figura 5.3: Variación de todos los componentes principales en función de los días de la semana.



El segundo componente o factor principal (PC2) explica un 16,07% de la varianza. Los parámetros con valores más altos son la $[CE]_{rb}$ (0,882), $[CE]_i$ (0,834), $[CE]_e$ (0,791) y $[Ta]_{rb}$ (0,590), con estos datos podríamos decir que el PC2 corresponde al componente inorgánico o iónico del proceso, es decir, la concentración de solutos en disolución, en la que hay que tener en cuenta que la temperatura modifica la conductividad, siendo esta mucho mayor en verano que en el resto de estaciones (**Figura 5.2**). Además, el aumento de las concentraciones de iones durante el periodo de verano podría explicarse por una evaporación más fácil del agua debido a un aumento de la temperatura y la irradiación del sol, significativamente diferente de las otras estaciones del año,

las cuales tienen mayores precipitaciones, lo que lleva a un efecto de dilución general. Resultados similares han sido reportados previamente en nuestra área para la evolución de los sólidos en los lodos de aguas residuales (Bayo *et al.*, 2015). El factor principal 2 (PC2) que representa, por tanto, el fondo inorgánico y geoquímico de nuestra agua, con el valor máximo de puntuación positiva para el lunes. Esto se debe a una disminución de la actividad antropogénica durante los fines de semana, que haría destacar este segundo componente sobre el resto por el fondo natural inorgánico de las aguas en la Región de Murcia. A medida que avanza la semana, El PC2 se reduce por el efecto de amortiguación de la carga orgánica en el agua residual, representada por el PC4. La eficiencia, que está representada por el componente 3 (PC3), muestra una tendencia opuesta a la evolución durante la semana del PC2, es decir, tiene valores muy bajos los lunes y va aumentando hasta alcanzar su punto más alto los jueves.

El tercer componente principal o factor (PC3) explica un 10,94% de la varianza. Los parámetros con los valores más altos son la [DQO]_e (0,776), [SS]_e turbidez (0,692) y [DBO₅]_e (0,601). Podríamos decir que el PC3 hace referencia a la eficiencia de la EDAR, puesto que las tres variables originales fundamentales que lo conforman son variables de salida de la EDAR. Se observa que las muestras recogidas en invierno, primavera y otoño de los tres años de estudio (2010 a 2012), dan valores mucho más altos que los del verano, que tiene valores muy negativos en este tercer factor, esto podría explicarse por la salida

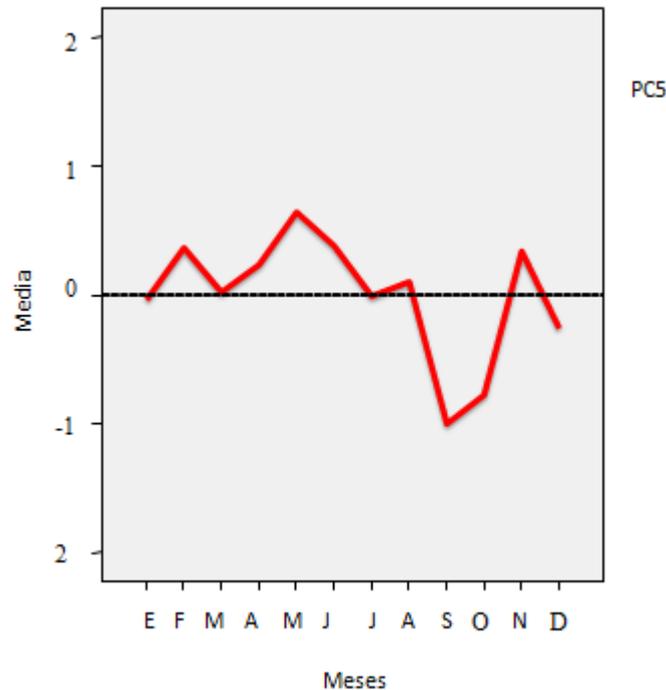
de los habitantes durante la temporada de verano con la correspondiente disminución de la actividad industrial de la zona, lo que mejora las variables de salida del agua de la EDAR.

El componente principal 3 (PC3) muestra los valores más altos para los martes, miércoles y jueves, y los valores más bajos los lunes y los viernes, justo después y antes del fin de semana, cuando la actividad antropogénica en la ciudad disminuye, afectando a la carga orgánica (PC4) que llega a la EDAR. En estos tres días se observa una disminución de la eficiencia (aumento del componente PC3), aunque siempre dentro de los valores establecidos por la legislación vigente.

El cuarto componente principal o factor (PC4) explica un 8,08% de la varianza total. Los parámetros que lo conforman son $[DQO]_i$ (-0,596) y $[DBO_5]_i$ (-0,581), lo que indica claramente que el PC4 correspondería con la carga de materia oxidable (orgánica e inorgánica) que llega con el agua residual a la EDAR. El hecho de que las puntuaciones sean negativas hace que un incremento en ambas variables de entrada se traduzca en una disminución del factor cuarto, y viceversa.

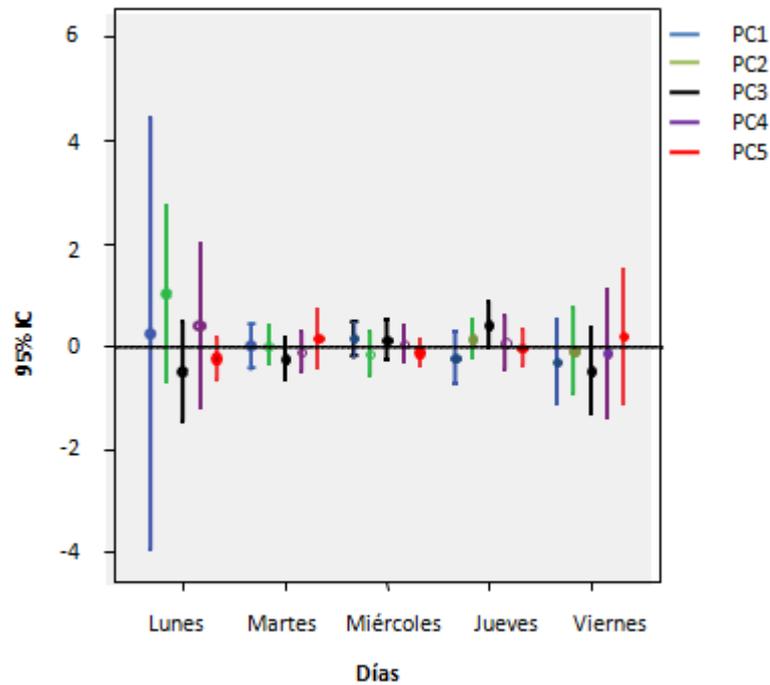
El quinto componente principal o factor (PC5) explica un 7,03% de la varianza. Los parámetros con puntuación más alta fueron los $[SS]_i$ (0,519) y $[N-NO_3]_i$

(0,468). Con estos resultados, podríamos pensar que el PC5 pudiera corresponderse con la recarga en la EDAR de nitratos incluidos en las partículas (sólidos en suspensión), por lo que este quinto componente podemos asociarlo con la erosión eólica del terreno. El viento constituye uno de los factores climáticos más importantes de la comarca, en las que predominan flujos de aire del tercer y primer cuadrante a lo largo del año, debido al efecto de barrera que suponen las sierras litorales de las cordilleras Béticas, que favorecen un cambio de rumbo a suroeste, y a la canalización de los flujos en el portillo tectónico que constituye Cartagena y el Mar Menor (Alarcón, A., 2014). En la **Figura 5.4**, podemos observar como el componente principal 5 varía a lo largo de los meses del año, fuertemente influenciado por el viento. Podemos ver un primer pico el mes de febrero, esto podría explicarse porque febrero es un mes muy tormentoso en Cartagena (González, J., 2006). En esta **Figura 5.4**, es de vital importancia valorar las brisas Tierra-Mar y Mar-Tierra, puesto que estas brisas ejercen erosión eólica contra el medio y transportan partículas en suspensión, estas brisas son más acusadas en las estaciones de calor y en días despejados sin nubes, por eso las estaciones de Mayo y Junio, presentan el pico más grande, puesto que estos meses en Cartagena marcan el inicio de la época más calurosa.

Figura 5.4: PC5 representado en función de los meses del año

A continuación, realizamos unas barras de error de los PCs, que nos permitan identificar la variabilidad de la media empleada como función de resumen del gráfico. La estructura del gráfico se basa en un punto central que identifica el valor de la media, la cual se ubica en una línea vertical; la longitud de la línea (barra de error) indica un intervalo de confianza (porcentaje de los datos) o un número especificado de errores típicos, o inclusive un número específico de desviaciones típicas. Las barras de error pueden extenderse en una o en ambas direcciones desde la media.

Figura 5.5: Barras de error de los PCs, en función de los días de la semana



A simple vista, podríamos decir que las barras de error muestran que no existen diferencias significativas entre los PCs, puesto que las barras de error en su gran parte, están solapadas. Para asegurarnos realizamos un análisis de la varianza ANOVA, en el que la hipótesis nula es que las medias no muestran diferencias estadísticamente significativas, y si la hipótesis alternativa que las medias son diferentes. En este ANOVA se realiza la prueba de homogeneidad de varianza, que podemos ver en las **Tablas 5.7 y 5.8**.

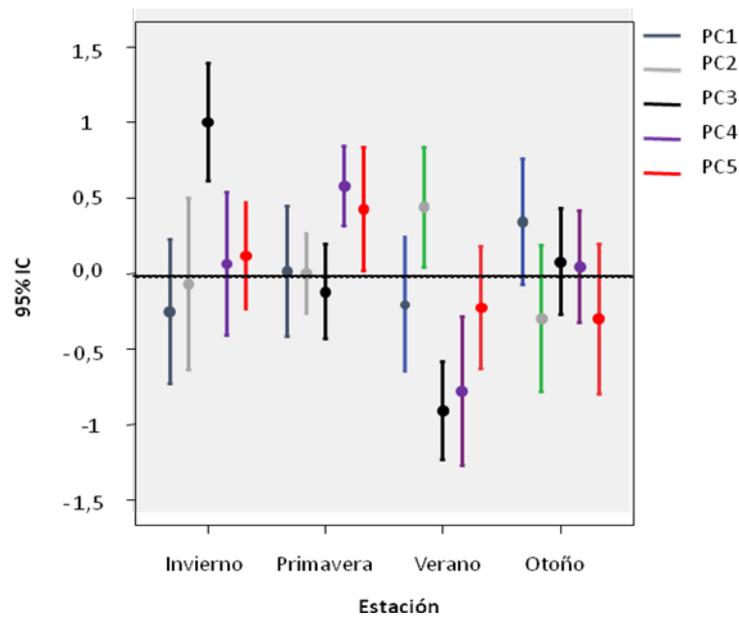


Figura 5.6: Barras de error de los PCs, en función de las estaciones del año.

Tabla 5.7: Prueba de homogeneidad de varianzas para los días de la semana

	Estadístico de Levene	gl1	gl2	Sig.
PC1	1,051	4	82	0,386
PC2	0,861	4	82	0,491
PC3	1,076	4	82	0,374
PC4	0,323	4	82	0,862
PC5	2,995	4	82	0,023

El nivel de significación es mayor a 0,05 en todos los casos, excepto para el PC5, por lo que asumimos que no hay diferencias estadísticamente significativas

entre los cuatro primeros PCs, pero sí para el quinto, a un nivel de significación del 95%.

Tabla 5.8: Prueba de homogeneidad de varianzas para las estaciones del año

	Estadístico de Levene	gl1	gl2	Sig.
PC1	0,486	3	83	0,693
PC2	2,909	3	83	0,039
PC3	0,922	3	83	0,434
PC4	1,864	3	83	0,142
PC5	0,294	3	83	0,829

El nivel de significación es mayor a 0,05 en todos los casos excepto para el PC2, por lo que asumimos que no hay diferencias estadísticamente significativas entre PC1, PC3, PC4 y PC5, pero si con el PC2. Este resultado es coincidente con lo descrito con anterioridad para este segundo componente, y remarca su carácter o representación del fondo geoquímico de nuestras aguas, que principalmente se manifiesta en la EDAR de la ciudad en verano, cuando disminuye la actividad antrópica e industrial en la ciudad de Cartagena.

Como el objetivo general del presente trabajo es conocer la eficiencia en la EDAR, hemos realizado unas barras de error para el PC3 (**Figura 5.8**),

correspondiente a la eficiencia, además de un ANOVA (**Tabla 5.10**) para saber si hay diferencias estadísticamente significativas.

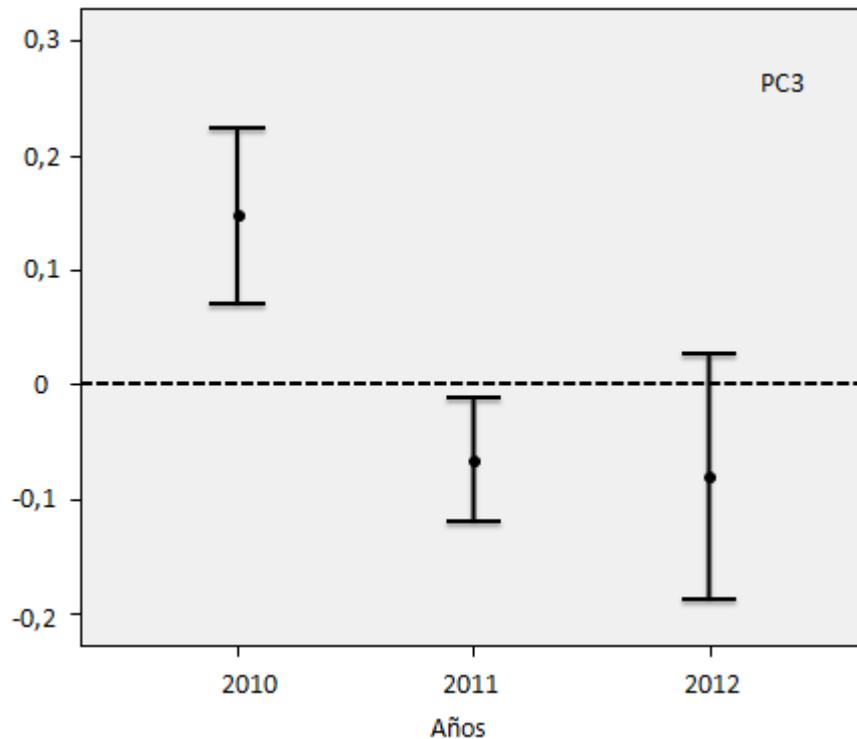


Figura 5.8: Barras de error del PC3 en función de los tres años de estudio.

En estas barras de error podemos observar cómo evoluciona el PC3 (eficiencia) en la EDAR, en los tres años de estudio.

El análisis de varianzas (ANOVA) indica que si hay diferencias estadísticas significativas para la eficiencia en los años de estudio, tal y como se observa en la **Tabla 5.9**.

Tabla 5.9: ANOVA del factor PC3 referido a los tres años de estudio.

	Suma de cuadrados	gl	Media cuadrática	F	Sig.
Inter-grupos	11,863	2	5,931	9,394	,000
Intra-grupos	690,137	1093	,631		
Total	702,000	1095			

En este caso, el valor del parámetro estadístico *F-Snedecor* y su significación estadística indican diferencias, al menos entre el primer año de estudio (2010) y los otros dos restantes (2011 y 2012).

Observamos una tendencia al descenso en el factor PC3 por año de estudio, lo que indicaría un valor inferior en sólidos en suspensión, DQO y DBO₅ en el agua residual efluente, es decir, una mejora en la eficiencia de funcionamiento de la depuradora. Estos resultados pueden deberse al buen trabajo realizado por la administración regional, que ha puesto en marcha campañas de sensibilización pública sobre la importancia de ahorrar agua y ha animado a la población a separar de forma adecuada los residuos generados en el domicilio y evitar su incorporación al sistema de saneamiento municipal, evitando el exceso de carga en el sistema de alcantarillado. Según lo indicado por Pietrzyk-Kaszyńska y Grodzińska-Jurczak (2015), el nivel administrativo regional

desempeña un papel central en la implementación de políticas que en muchos casos buscan unos objetivos de carácter más internacional. Por otra parte, en octubre de 2010, se puso en marcha un tanque de tormentas con un volumen total de 10.000 m³, localizado antes de las rejillas de desbaste de la depuradora. Este tanque se empieza a llenar cada vez que se alcanza un caudal de entrada de más de 1.200 m³/h, lo que proporciona un sistema óptimo para amortiguar las cargas orgánicas puntuales del agua de entrada y promueve una composición más homogénea del influente, que hace que el agua depurada cumpla con la Directiva 91/271/CEE.

5.2 Análisis de conglomerados jerárquico

Se llevó a cabo un análisis de agrupamiento jerárquico de exploración (HCA) con los 23 parámetros de calidad del agua que resultaron ser significativo en el análisis del componente principal, usando el método de Ward y la distancia euclidiana como una medida de disimilitud. En nuestro estudio, el análisis de conglomerados se ha utilizado para determinar asociaciones de grupos y para evaluar la afinidad entre las variables, tal y como han realizado con anterioridad Hussain *et al.* (2008).

El dendrograma que obtenemos se puede dividir en dos grupos, A y B, como podemos apreciar en la **Figura 5.9**. Estos dos grupos están separados a una

distancia de unión inferior a 25 y representa dos calidades de agua diferentes. El primer grupo (A) estaría formado por parámetros de calidad del agua del afluente, mientras que el segundo grupo (B) incluiría la mayor parte de los parámetros de calidad del efluente, que corresponde a la eficiencia de funcionamiento de la EDAR. Lourenço *et al.* (2006) encontraron un resultado similar en una planta de tratamiento de aguas residuales incluida en una gasolinera, con dos grupos bien diferenciados, uno que representaba las muestras de agua residual y el otro que incluía el agua del tanque de fuego y el agua del grifo, en definitiva, agua de mejor calidad.

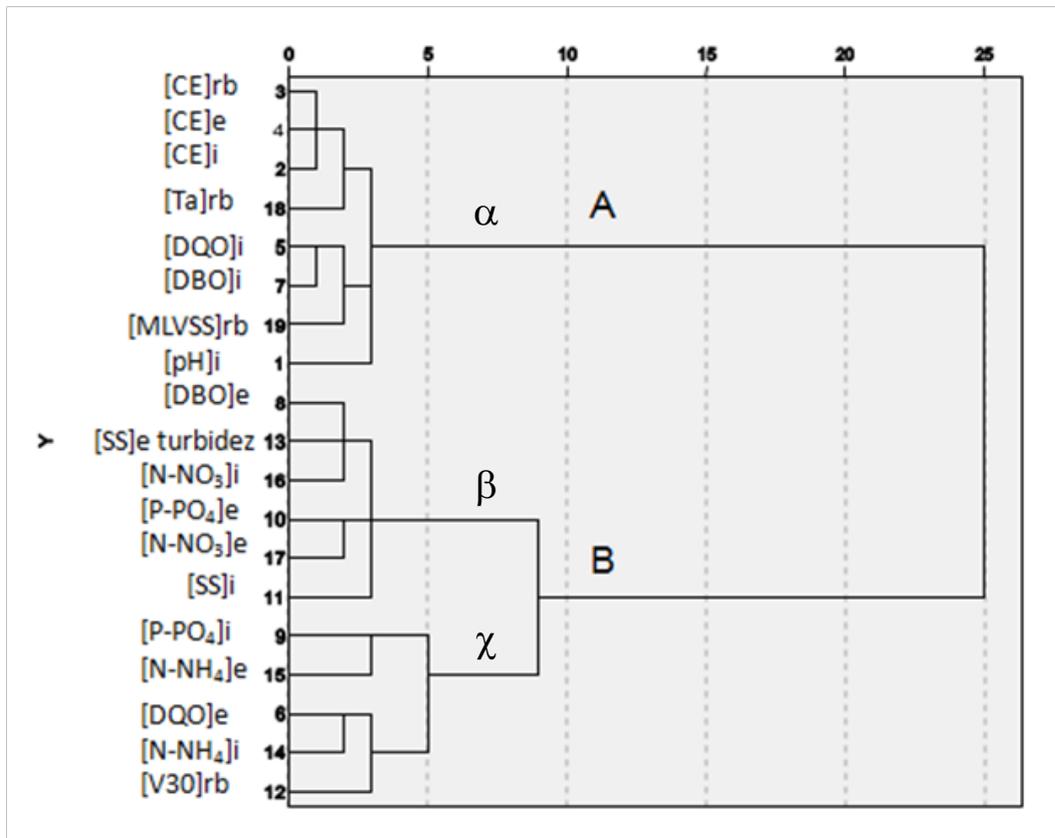


Figura 5.9: Dendrograma con vinculación Ward

En un segundo paso, y utilizando un valor de criterio de distancia de reescalado entre 5 y 10, encontramos tres grupos bien diferenciados: un primer grupo α representado por los mismos parámetros de entrada previamente identificados en el grupo A, un segundo grupo β , que principalmente recoge la salida de carga orgánica y nutrientes de la depuradora, y un tercer grupo χ que representa la salida de DQO y entrada y salida de amonio.

Cuando el dendrograma resultante fue interpretada por las distancias de enlace más cortas (<5), los resultados fueron similares a los reportados para el análisis de componentes principales, con cinco grupos principales. Con el fin de probar las diferencias de un grupo a otro, y para revelar las tendencias de los principales grupos, los valores de los diferentes parámetros fueron autoescalados dividiendo el valor medio del parámetro por el promedio máximo para cada parámetro de calidad del agua, excluyendo de este estudio el PC5, por ser el que menos varianza total explicaba (7,03%). Así, la escala del eje “y” fue del 0 al 1, tal como se presenta en la **Figura 5.10**.

El Grupo 1 mostró sus valores máximos para el pH y las distintas variables de entrada a la EDAR, es decir, $[\text{NH}_4\text{-N}]_i$, $[\text{COD}]_i$, $[\text{DBO}]_i$, $[\text{SS}]_i$ y $[\text{PO}_4\text{-P}]_i$, similar al cuarto componente encontrado en el análisis de componentes principales. Además, las mayores diferencias en este primer grupo fueron para

los valores autoescalados de [COD]_e, [PO₄-P]_e y [SS]_e, lo que indicaría que el análisis de conglomerados jerárquicos podría separar de manera eficiente entre las aguas residuales crudas y aguas residuales tratadas.

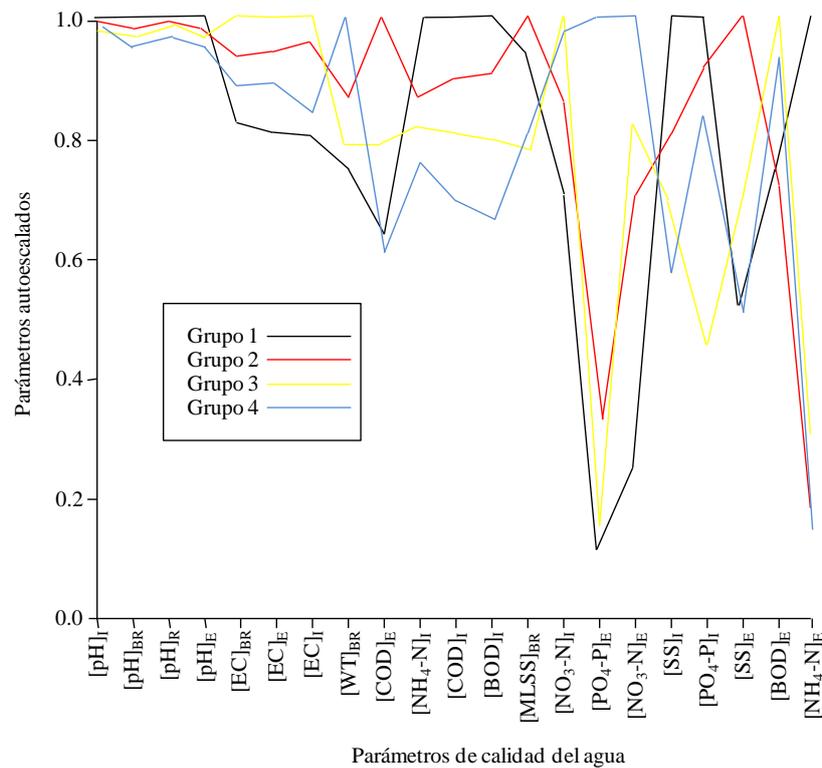


Figura 5.10: Clasificación de los resultados del clúster según valores autoescalados para los diferentes parámetros de calidad del agua

El segundo grupo (Grupo 2) anotó los más altos valores para [COD]_e y [SS]_e, representando ambos la descarga inorgánica de la EDAR, similar a la calidad de los efluentes representado por PC3. El Grupo 3 incluyó, principalmente, la

conductividad eléctrica del agua, y podría ser identificado como el componente geoquímico del agua en nuestra región, que contiene típicamente sales de concentraciones variables de origen pedogénico (salinidad primaria), tal y como han señalado con anterioridad por Acosta *et al.* (2011). Este tercer grupo sería similar al representada por PC2 en el análisis de componentes principales.

El último grupo (Grupo 4) incluye parámetros relacionados con nutrientes en las aguas residuales tratadas; es decir, $[\text{PO}_4\text{-P}]_e$ y $[\text{NO}_3]_e$, algo similar a PC1 en el análisis de componentes principales. De hecho, es importante señalar que la diferencia en los valores autoescalados de $[\text{PO}_4\text{-P}]_e$ y $[\text{NO}_3]_e$ para este cuarto grupo y el resto es muy alta, alcanzando los coeficientes de variabilidad observados más alto; cerca del 90% para $[\text{PO}_4\text{-P}]_e$ y 75% para $[\text{NO}_3]_e$, con una distancia máxima al Grupo 1. Por el contrario, por ejemplo, los valores de los coeficientes de variabilidad para el pH no superaban el 4% en ningún caso.

6 Conclusiones

De la aplicación de dos técnicas quimiométricas diferentes: análisis de componentes principales (PCA) y análisis de conglomerados jerárquicos (HCA), para el estudio de la estructura inherente a un conjunto de datos de calidad del agua de la EDAR urbana Cabezo Beaza, para los años 2010, 2011 y 2012, podemos extraer las siguientes conclusiones:

- 1) Para el análisis de componentes principales se extrajeron un total de siete factores o componentes que representaban el 78,73% de la varianza total de los datos, aunque los dos últimos factores fueron descartados por contener un solo parámetro ya representado en algún otro de los componentes.
- 2) Esta pérdida supuso una disminución del 10,43% de la varianza total explicada, quedando esta en un 68,30%, de forma similar a los resultados obtenidos por otros autores en estudios similares (Vialle *et al.*, 2011; Zhang *et al.*, 2010; Gazzaz *et al.*, 2012).
- 3) El primer factor (PC1) explicó la circulación de nutrientes y de materia orgánica dentro de la EDAR, tanto para las aguas residuales tratadas como no tratadas, demostrando ser mayor durante las estaciones de

primavera y otoño, cuando se aplican al suelo fertilizantes y residuos del ganado.

- 4) El segundo componente (PC2) explicaba un 16,07% de la varianza total, y representaba el componente inorgánico y geoquímico del agua. Este componente demostró aumentar durante la estación de verano, cuando la ciudad se despuebla y la producción industrial disminuye, quedando por encima este componente natural del agua, que en otras estaciones queda enmascarado por la propia carga orgánica.
- 5) Así mismo, este PC2 mostró los valores más altos para el primer día de la semana, debido a una disminución de la actividad antropogénica e industrial durante los fines de semana, que haría destacar este segundo componente sobre el resto.
- 6) El tercer componente principal (PC3) explica un 10,94% de la varianza total, y representa la eficiencia de depuración de la EDAR. Este componente también presentó una variación estacional, siendo más óptimo para la estación más cálida.
- 7) La evolución semanal del tercer componente (PC3) fue opuesta a la del PC2, presentando los valores más bajos (mayor eficiencia) los días cercanos al fin de semana (lunes y viernes).

- 8) El cuarto componente del PCA (PC4) explicaba un 8,08% de la varianza total, y representa la carga inorgánica y orgánica que llega a la depuradora.
- 9) El quinto componente (PC5) explicaba un 7.03% de la varianza total, y puede estar relacionado con la carga de nitrógeno en la depuradora debida a la erosión eólica.
- 10) El análisis de conglomerados jerárquico desveló, fundamentalmente, una diferencia entre parámetros de entrada y de salida a la depuradora, aunque un análisis más detallado permitió establecer grupos muy similares a los obtenidos por análisis de componentes principales.

7. Referencias bibliográficas

- Acosta, J.A., Faz, A., Jansen, B., Kalbitz, K., Martínez-Martínez, S. 2011. Assessment of salinity status in intensively cultivated soils under semiarid climate, Murcia, SE Spain. *J. Arid. Environ.* 75, 1056–1066.
- Aguado, D., Rosen, C. 2008. Multivariate statistical monitoring of continuous wastewater treatment plants. *Eng. Appl. Artif. Intel.* 21, 1080–1091.
- Alarcón, V. Análisis del sistema de gestión de los recursos hídricos en el Municipio de Cartagena (Región de Murcia. Congreso nacional de medio ambiente (CONAMA 2014).
- APHA, Standard Methods for the Examination of Water and Wastewater. 2012. 22nd ed., in: E.W. Rice, R.D. Baird, A.D. Eaton, L.S. Clesceri (Eds.). American Public Health Association, American Water Works Association & Water Environment Federation, Washington D.C.
- Astel A., Tsakovski S., Simeonov V., Reisenhofer E., Piselli S., Barbieri P. 2008. Multivariate classification and modeling in surface water pollution estimation. *Anal Bioanal Chem.* 390, 1283-1292.
- Astel, A., Michalski, R., Lyko, A., Jablonska-Czapla, M., Bigus, K., Szopa, S., Kwiecinska, A. 2014. Characterization of bottled mineral

- waters marketed in Poland using hierarchical cluster analysis. *J. Geochem. Explor.* 143, 136–145.
- Baker, J. 2010. A cluster analysis of long range air transport pathways and associated pollutant concentrations within the UK. *Atmospheric Environment.* 44, 563-571.
- Bayo, J., López-Castellanos, J., Martínez-García, R., Alcolea, A., Lardín, C. 2015. Hydrocyclone as a cleaning device for anaerobic sludge digesters in a wastewater treatment plant. *J. Clean. Prod.* 87, 550–557.
- Beamonte, E.; Bermúdez, J.; Casino, A.; Veres, E. (2004). Un indicador global para la calidad del agua. Aplicación a las aguas superficiales de la Comunidad Valenciana, en *Revista Estadística Española*, Volumen 46, número 156; pp. 189-204.
- Beamonte, E.; Bermúdez, J.; Casino, A.; Veres, E. (2007). A statistical study of the quality of surface water intended for human consumption near Valencia (Spain), en *Journal of Environmental Management*, número 83; pp. 307-314.
- Beamonte, E.; Bermúdez, J.; Casino, A.; Veres, E. (2010). Water quality indicators: Comparison of a probabilistic index and general quality index. The case of the Confederación Hidrográfica del Júcar (Spain), en *Ecological Indicators*, número 10; pp.1049 -1054.

- Beamonte, E.; Bermúdez, J.; Casino, A.; Veres, E. (2012). Análisis de la calidad general del agua superficial en la cuenca hidrográfica del Júcar: periodo 2000-2009, en *M+A. Revista Electrónica de Medioambiente*, número 12; pp. 18-32.
- Beamonte, E.; Casino, A.; Veres, E. J. (2004). La calidad del agua en ciertas estaciones de control del canal Júcar-Turia (período 1994-2001), en *Revista Española de estudios Agrosociales y Pesqueros*, número 201; pp. 105-126.
- Beirle, S., Platt, U., Wenig, M., Wagner, T. 2003. Weekly cycle of NO₂ by GOME measurements: a signature of anthropogenic sources. *Atmos. Chem. Phys.* 3, 2225–2232.
- Bingöl, D., Ay, Ü., Bozbas, S.K., Uzgören, N. 2013. Chemometric evaluation of the heavy metals distribution in waters from the Dilovasi region in Kocaeli, Turkey. *Mar. Pollut. Bull.* 68, 134–139.
- Chen, H., Lu, X., Li, L. Y., Gao, T., Chang, Y. 2014. Metal contamination in campus dust of Xi'an, China: A study based on multivariate statistics and spatial distribution. *Sci. Total Environ.* 484, 27–35.
- Council Directive 91/271/EEC of 21 May 1991 concerning urban waste water treatment (amended by Directive 98/15/EC). *Off J Eur Union* L135/40 (30 of May 1991).
- Cuadras, C. M. 1981. *Métodos de Análisis Multivariante*. Eunibar,

Barcelona.

Cuadras, C.M. 1981. Análisis y representación multidimensional de la variabilidad. *Intem. Symp. Concept. Meth. Paleo.* 287-297.

Barcelona.

Cuadras, C.M. 2014. Nuevos métodos de análisis multivariante.

Barcelona.

Gazzaz, N.M., Yusoff, M.K., Ramli, M.F., Aris, A.Z., Juahir, H. 2012.

Characterization of spatial patterns in river water quality using chemometric pattern recognition techniques. *Mar. Pollut. Bull.* 64, 688–698.

González, J. *Climatología de tormentas en España.*2006.

Hernández, F.; Molinos, M.; Sala, R. (2009). Valoración Económica de los Beneficios Ambientales del Proceso de Depuración de Aguas Residuales en Rect@, Volumen 17.

Hussain, M., Ahmed, S.M., Abderrahman, W. 2008. Cluster analysis and quality assessment of logged water at an irrigation project, eastern Saudi Arabia. *J. Environ. Manage.* 86, 297–307.

Johar, M., Savage, E. 2013. Discovering unhealthiness: evidence from cluster analysis. *Annals of Epidemiology.* 23, 614-619

Kaiser, H.F. 1960. The application of electronic computers to factor analysis. *Educ. Psychol. Meas.* 20, 141–151.

- Kaiser, H.F. 1974. An index of factorial simplicity. *Psychometrika* 39, 31–36.
- Karavoltsos, S.; Sakellari, A.; Mihopoulos, N.; Dassenakis, M.; Scoullou, M.J. (2008). Evaluation of the quality of drinking water in regions of Greece, en *Desalination*, número 224: pp. 317-329
- Kaufman, L., Rousseeuw, P.J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.
- Kiralj, R., Ferreira, M.M.C. 2006. The past, present, and future of chemometrics worldwide: some etymological, linguistic and bibliometric investigations. *J. Chemometrics*. 20, 274-272.
- Lourenço, N.D., Chaves, C.L., Novais, J., Menezes, J.C., Pinheiro, H.M., Diniz, D. 2006. UV spectra analysis for water quality monitoring in a fuel park wastewater treatment plant. *Chemosphere* 65, 786–791.
- Lu, L., Li, X., Gao, J. 2011. Evaluation of Aging Behavior of Medium Density Polyethylene in Natural Environment by Principal Component Analysis. *Journal of Applied Polymer Science*. 125, 19-23.
- Naughton, O., Hynds, P.D. 2014. Public awareness, behaviours and attitudes towards domestic wastewater treatment systems in the Republic of Ireland. *J. Hydrol.* 518, 108–119.
- Noori, R., Karbassi, A., Sabahi, M.S. 2010. Evaluation of PCA and

- Gamma test techniques on ANN operation for weekly solid waste prediction. *J. Environ. Manage.* 91, 767–771.
- Olsen, R.L., Chappell, R W., Loftis, J.C. 2012. Water quality sample collection, data treatment and results presentation for principal components analysis–literature review and Illinois River watershed case study. *Water Res.* 46, 3110–3122.
- Ouali, A., Azri, C., Medhioub, K., Ghrabi, A. 2009. Descriptive and multivariate analysis of the physico-chemical and biological parameters of Sfax wastewater treatment plant. *Desalination* 246, 496–505.
- Peng, X., Li, X., Shi, X., Guo, S. 2014. Evaluation of the aroma quality of Chinese traditional soy paste during storage based on principal component analysis. *Food chemistry.* 151, 532-538.
- Pietrzyk-Kaszyńska, A., Grodzińska-Jurczak, M. 2015. Bottom-up perspectives on nature conservation systems: The differences between regional and local administrations. *Environ. Sci. Policy* 48, 20–31.
- Platikanov, S., Rodriguez-Mozaz, S., Huerta, B., Barceló, D., Cros, J., Batle, M., Poch, G., Tauler, R. 2014. Chemometrics quality assessment of wastewater treatment plant effluents using physicochemical parameters and UV absorption measurements. *J.*

- Environ. Manage. 140, 33–44.
- Reimann, C., Filzmoser, P., Garrett, R.G. 2002. Factor analysis applied to regional geochemical data: problems and possibilities. *Appl. Geochem.* 17, 185–206.
- Rescala, C., Devincenzi, G., Rohde, G., Bonaffini M. L., Giraud, M. V., Bernaola, G. 2012. Two models to determine the efficiency of a construction firm.
- Rodriguez, B., Peña-Mendez, E., Havel, J., Rodriguez E. M., Romero, C. 2010. Cluster analysis and artificial neural networks multivariate classification of onion varieties. *J. Agric. Food Chem.* 58, 11435-11440.
- Saba, T., Su, S. 2013. Tracking polychlorinated biphenyls (PCBs) congener patterns in Newbark Bay surface sediments using principal component analysis (PCA) and positive matrix factorization (PMF). *J. Hazard. Mater.* 260, 634–643.
- Salvati, L., Carlucci, M. 2014. Zero Net Land Degradation in Italy: The role of socioeconomic and agro-forest factors. *J. Environ. Manage.* 145, 299–306.
- Saz, S.; Hernández, F.; Sala, R. (2009). Estimación del valor económico de la calidad del agua de un río mediante una doble aproximación: una aplicación de los principios económicos de la Directiva Marco

- del Agua, en *Economía Agraria y Recursos Naturales*”, Volumen 9, número 1; pp. 37-63.
- Schipa, I., Tanzarella, A., Mangia, C. 2009. Differences between weekend and weekday ozone levels over rural and urban sites in Southern Italy. *Environ. Monit. Assess.* 156, 509–523.
- Sevilla, M.; Torregrosa, T. (2009-2011). Alternativas hídricas y agricultura. Documento de trabajo del Instituto Interuniversitario de Economía Internacional (IEI). Alicante.
- Sevilla, M.; Torregrosa, T.; Moreno, L. (2010). Un panorama sobre la economía del agua, en *Estudios de Economía Aplicada*, Volumen 28, número 2; pp. 265-304.
- Shrestha, S.; Kazama, F. (2007). Assesment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan, en *Enviroment Modelling & Software*, número 22; pp. 464-475.
- Singh, K.P., Malik, A., Mohan, D., Sinha, S., Singh, V K. 2005. Chemometric data analysis of pollutants in wastewater: a case study. *Anal. Chim. Acta* 532, 15–25.
- Sotelo, M. (2013). Territorio y medio ambiente en la Comunidad de Madrid. Las infraestructuras históricas, nuevos paisajes culturales del agua, en *M+A. Revista Electrónica de Medioambiente*,

Volumen 14, número 1; pp. 87-115.

Storline, M.R. 1981. The status of multiple comparisons: Simultaneous estimation of all pairwise comparisons in one-way ANOVA designs. *Am. Stat.* 35, 134–141.

Tchobanoglous, V.G., Burton, F.L., David, H., 2003. *Wastewater Engineering: treatment and Reuse*, 4th ed., McGraw-Hill: New York, USA.

Terrádez, M. 2003. Análisis de componentes principales. Proyecto e-Math. www.uoc.edu.

Vera, J.F. (2006). Agua y modelo de desarrollo turístico: la necesidad de nuevos criterios para la gestión de los recursos, en *Boletín de la Asociación de Geógrafos Españoles (A.E.E.)* número 42; pp. 155-178.

Vialle, C., Sablayrolles, C., Lovera, M., Jacob, S., Huau, M.C., Montrejaud-Vignoles, M. 2011. Monitoring of water quality from roof runoff: Interpretation using multivariate analysis. *Water Res.* 45, 3765–3775.

Wunderlin, D.A., Díaz, M.D., Amé, M.V., Pesce, S.F., Hued, A.C., Bistoni, M.A. 2001. Pattern recognition techniques for the evaluation of spatial and temporal variations in water quality . A case study: Suquía river basin (Córdoba-Argentina). *Water Res.* 35,

2881–2894.

Xiaorui, Z. General Guidelines for Methodologies on Research and Evaluation of Traditional Medicine. 2000. Geneva. WHO/EDM/TRM/2000.1.

Zhang, W., Zhang, Y. Taniyasu, S., Yeung, L.W.Y., Lam, P.K.S., Wang, J., Li, X., Yamashita, N., Dai, J. 2013. Distribution and fate of perfluoroalkyl substances in municipal wastewater treatment plants in economically developed areas of China. *Environ. Pollut.* 176, 10–17.

Zhang, Z., Tao, F., Du, J., Shi, P., Yu, D., Meng, Y., Sun, Y. 2010. Surface water quality and its control in a river with intensive human impacts—a case study of the Xiangjiang River, China. *J. Environ. Manage.* 91, 2483–2490.

Zibert, J., Praznikar, J. 2012. Cluster analysis of particulate matter (PM10) and black carbon (BC) concentrations. *Atmospheric Environment.* 57, 1-12.

Zuthi, M.F.R., Guo, W.S., Ngo, H.H., Nghiem, L.D., Hai, F.I., 2013. Review: Enhanced biological phosphorus removal and its modeling for the activated sludge and membrane bioreactor processes. *Bioresource Technol.* 139, 363–374.